

변 혜 원 교수지도  
석사학위 청구논문

# 사운드의 실시간 비주얼 시뮬레이션

2013

성신여자대학교 대학원

컴퓨터학과

권경아

# 사운드의 실시간 비주얼 시뮬레이션

변 혜 원 교수지도

이 논문을 석사학위논문으로 제출함

2013년 5월

성신여자대학교 대학원

컴퓨터학과

권경아

# 인 준 서

권경아의 석사학위 논문으로 인준함.

심사위원 홍 승 필 인

심사위원 김 태 훈 인

심사위원 변 혜 원 인

성신여자대학교 대학원

## 논문개요

최근 과학 기술의 발전과 함께 서로 다른 감각기관에 바탕을 둔 예술을 융합시킨 새로운 형태의 연구가 활발히 진행되고 있다. 특히 시각과 청각을 연관 지으려는 시도는 사운드 시각화라는 개념으로 그 범위를 넓혀가고 있다. 사운드 시각화는 청각적 요소와 시각적 요소를 연동하는 것으로 음악을 시각적으로 표현하는 것을 말한다. 이러한 사운드 시각화는 우리 주변에서 쉽게 찾아 볼 수 있는데 음악에 맞춰 움직이는 음악 분수나 음악에 따라 다양한 시각효과를 만들어내는 미디어 플레이어가 대표적인 예라고 할 수 있다. 음악에 따른 시각효과로는 이퀄라이저 효과가 일반적으로 사용된다. 이퀄라이저 효과는 사운드의 템포, 데시벨 등의 특징을 시각적으로 표현한 것으로 막대그래프 형태가 일반적이다. 막대그래프 외에도 파동, 파티클 등의 다양한 형태가 존재하지만 대부분 그래픽 패턴과 사운드 주파수에 따른 색상 변화에만 주목하고 있다.

본 논문에서는 사운드 이퀄라이저 기법을 활용하여 이미지가 가지고 있는 콘텐츠의 형태를 유지하는 2.5차원의 이미지 이퀄라이저를 제안한다. 여기서 콘텐츠란 이미지가 가지고 있는 텍스트, 캐릭터 등이 될 수 있으며, 이러한 콘텐츠를 유지시키기 위한 방법으로 마스킹 기법을 이용한다. 동적인 시뮬레이션을 위해서는 사운드가 활용된다. 입력되는 사운드를 기반으로 이퀄라이저를 생성하기 위해서는 사운드 분석과정이 필요하다. 제안하는 시스템은 사운드를 분석하기 위해 고속 푸리에 변환(Fast Fourier Transform : FFT) 알고리즘을 이용한다. 고속 푸리에 변환을 이용해 사운드의 주파수 성분을 추출하고, 추출된 데이터를 기반으로 이퀄라이저를 생성한다. 보통 그래픽 이퀄라이저는 2차원 공간상에 표현되는데 2차원 공간은 Degree of

Freedom(DOF)의 제약이 있어 입체적인 움직임을 표현하기에 어려움이 있다. 이러한 공간적 제약을 해결하기 위해 우리는 격자 기반으로 이퀄라이저를 생성한다. 이퀄라이저는 사운드의 주파수 성분에 의해 생성되는데 샘플링 된 주파수 데이터는 관심영역의 픽셀 수 보다 적어 주파수 값과 관심영역 상의 픽셀을 1:1 매칭 시킬 수 없다. 제한된 수의 주파수 데이터를 관심영역과 매칭 시키기 위해 모듈러 기법을 사용하고 생성된 이퀄라이저에 가우시안 필터(Gaussian Filter)를 적용하여 컨볼루션 시킴으로써 음악과 연동되는 이미지 이퀄라이저를 생성할 수 있다.

본 논문은 사용자가 원하는 이미지 내의 콘텐츠 형태를 유지시키면서 입력 음악과 동기화되는 이퀄라이저를 생성하여, 최종적으로는 정적인 2차원 이미지를 2.5차원의 동적 시뮬레이션으로 표현하였다. 이는 사용자의 피드백을 받아 이퀄라이저를 생성할 수 있다는 점에서 기존 이퀄라이저와 차별성이 있으며, 그래픽의 패턴에만 집중하는 기존 방식의 이퀄라이저에서 벗어난 새로운 형태의 이미지 이퀄라이저로써 음악과 함께 감상하는 그림 등에 유용하게 활용될 수 있다.

# 목 차

## 논문개요

I. 서론 .....	1
1. 연구배경 및 목표 .....	1
2. 논문의 독창성 .....	4
3. 논문구성 .....	5
II. 관련연구 .....	6
1. 사운드 시각화 .....	6
2. 오디오 시각화 작품 분석 .....	12
III. 시스템 구성 .....	16
IV. 사운드의 실시간 비주얼 시뮬레이션 .....	20
1. 이미지와 사운드 분석 .....	20
1.1. 이미지 관심 영역 지정 .....	20
1.2. 사운드 주파수 분석 .....	22
2. 격자 기반 시뮬레이션 .....	27
2.1. 이미지의 격자 맵핑 .....	27
2.2. 높이 맵 기반 주파수 맵핑 .....	28
2.3. 컨볼루션 .....	31
V. 실험결과 .....	33
VI. 결론 및 향후 연구 .....	43

## 참고문헌

Abstract(영문초록)

## 그림 목 차

[그림 1] 음악 분수와 미디어 플레이어 .....	2
[그림 2] 그래픽 이퀄라이저의 다양한 시각효과 .....	2
[그림 3] 본 논문에서 제안하는 사운드의 비주얼 시뮬레이션 .....	4
[그림 4] 가청 주파수와 가시 주파수 매칭을 통한 음-색 대응 .....	7
[그림 5] HSI 컬러 히스토그램을 이용한 색-음 변환 .....	8
[그림 6] 사운드에 의해 생성된 지형 .....	8
[그림 7] Sound Bound 결과 .....	9
[그림 8] 사운드 그래픽 이퀄라이저 결과 .....	10
[그림 9] 가야금 연주 인터페이스(좌)와 시각화 결과(우) .....	10
[그림 10] Creating Musical Fountain Shows 결과 .....	11
[그림 11] Harmonic Paper 인터페이스와 구조 .....	11
[그림 12] <Piano: as image media> 1995. Toshio Iwai .....	12
[그림 13] Rhyme Modulation - nong(弄) 2006. 오창근 .....	13
[그림 14] 공명의 Deep Sea, 2008 .....	13
[그림 15] 허클베리핀의 옐로우 콘서트, 2007 .....	14

[그림 16] Spacequatica 오스트리아 공연, 2008 .....	15
[그림 17] Picturedisko 2006. Mario Klingemann .....	15
[그림 18] 시스템 개요 .....	17
[그림 19] 마스킹 과정 .....	21
[그림 20] 마스킹을 통한 관심영역 선택 과정 .....	22
[그림 21] 시간 도메인과 주파수 도메인에서의 음향 신호 .....	23
[그림 22] 주파수의 주기 .....	23
[그림 23] 주파수 파형에 따른 소리의 특성 .....	24
[그림 24] 주파수 표현 .....	24
[그림 25] 푸리에 변환 .....	25
[그림 26] 입력 사운드의 스펙트럼 .....	26
[그림 27] 2차원 공간상의 이미지(좌) Degree of Freedom(우) .....	27
[그림 28] 이미지의 격자 맵핑 .....	28
[그림 29] 주파수 데이터와 관심영역 데이터의 불일치 .....	29
[그림 30] 주파수 데이터와 관심영역 내 픽셀 맵핑의 문제점 .....	29
[그림 31] 모듈러 방식을 이용한 주파수 데이터 - 관심영역 맵핑 .....	30

[그림 32] 관심영역의 형태적 특징을 유지하는 이퀄라이저 생성 .....	31
[그림 33] 일정한 패턴을 가지는 주파수 데이터의 문제점 .....	31
[그림 34] 가우시안 함수 그래프 .....	32
[그림 35] 5x5 가우시안 마스크 .....	32
[그림 36] 가우시안 필터를 사용한 컨볼루션 .....	32
[그림 37] 마우스 캐릭터의 소스 이미지(좌)와 마스크 이미지(우) .....	33
[그림 38] 주파수 값 분석을 위한 스펙트럼 결과 .....	35
[그림 39] 마우스 캐릭터의 시뮬레이션 결과 .....	36
[그림 40] 음표 심볼의 소스 이미지와 마스크 이미지 .....	37
[그림 41] 그림 40의 주파수 값 분석을 위한 스펙트럼 결과 .....	38
[그림 42] 음표 이미지의 시뮬레이션 결과 .....	39
[그림 43] 나뭇잎 소스 이미지(좌)와 마스크 이미지(우) .....	39
[그림 44] 나뭇잎 시뮬레이션의 정면(좌)과 측면(우) 결과 .....	40
[그림 45] 스머프 소스 이미지(좌)와 마스크 이미지(우) .....	40
[그림 46] 스머프 시뮬레이션 결과 .....	41
[그림 47] 입력 사운드의 주파수 값 변화 .....	41

[그림 48] 이미지에 맵핑된 주파수 값 변화 .....	42
[그림 49] 그림 48의 부분 확대 영상 .....	42

# I. 서론

## 1. 연구 배경 및 목표

최근 컴퓨터 기술의 발달과 함께 다양한 예술분야의 경계를 허물고, 새로운 형태의 융합예술을 위한 연구가 활발히 진행되고 있다. 음악을 시각적으로 표현하는 사운드 시각화도 이에 포함된다. 우리는 일반적으로 귀를 통해 음악을 들으며 즐거움을 느끼고, 눈을 통해 그림을 보고 감상한다. 하지만 미디어 시대가 도래 하면서 눈으로 듣는 듯한, 귀로 보는 듯한 감정을 느끼려는 다양한 시도들이 활발히 진행되고 있다. 다시 말해, 청각과 시각을 융합시켜 ‘공감각(synesthesia)’의 감정을 직접 실현시키고 느껴보고자 하는 것이다. 공감각이란 하나의 감각이 다른 영역의 감각을 자극하는 감각의 전이현상으로, 두 가지 감각을 함께 지각하는 경우를 말한다. 이러한 공감각을 느끼기 위한 방법 중 하나가 바로 청각과 시각을 연관 짓는 사운드 시각화이다. 청각과 시각을 연결하려는 시도는 현대에 들어서 새롭게 연구되는 분야는 아니며 디지털 기술이 존재하기 이전부터 고대 철학자, 사상가 및 예술가들에 의해 연구되어 왔다. 일찍이 피타고라스는 공진음을 수학적으로 분할하여 12음계를 발견해내고 각 음에 색채를 대응시키는 ‘색음이론’을 발표하였고, 물리학자인 뉴턴과 철학자 피테가 피타고라스의 이론을 뒷받침하며 시각적 요소와 청각적 요소를 연결하는 시도를 하였다. 이와 같은 사운드 시각화는 디지털 기술의 발달과 함께 미디어 아트, 전시, 공연 등의 분야에서 그 영역을 넓혀가고 있다. 우리 주변에서도 사운드 시각화를 쉽게 찾아볼 수 있다. 음악 소리에 맞춰 모양과 분위기를 변경하는 음악분수나

음악에 따라 움직이는 파형을 그리는 미디어플레이어가 그 예이다(그림1).

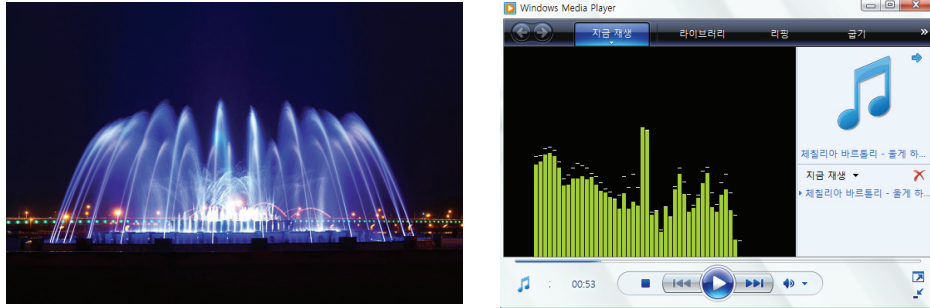


그림 1. 음악 분수와 미디어 플레이어

특히 미디어플레이어에서는 ‘시각화’ 라는 이름으로 사운드 이퀄라이저 효과를 제공한다. 음악의 파형을 표시하는 이퀄라이저에는 다양한 시각효과가 존재하지만 그림 2에서 보이는 바와 같이 음악에 따른 그래픽 패턴이나 주파수에 따른 색상 변화에만 집중하고 있다. 이는 사운드에 따라 변화하는 단순 반복적인 그래픽 효과로서 사용자가 지루함을 느낄 가능성을 가지고 있다.

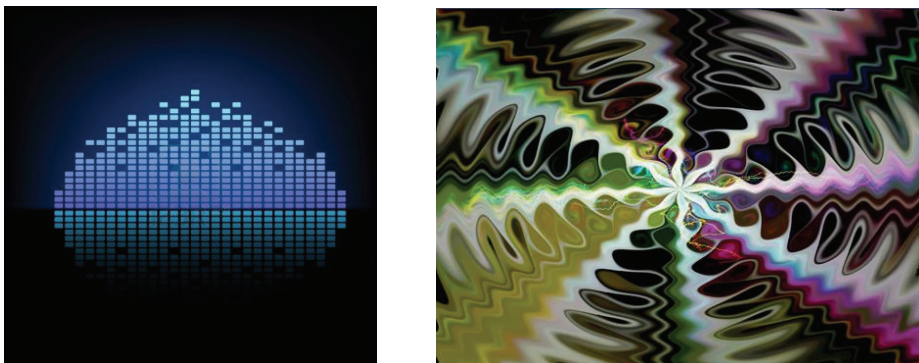


그림 2. 그래픽 이퀄라이저의 다양한 시각효과

본 논문에서는 사용자의 피드백을 받아 음악과 연동되는 이미지 이퀄라이저를 제안한다. 기존의 이퀄라이저와는 다르게 이미지를 소스로 하고, 이미지 내에 있는 콘텐츠의 형태적 특징을 유지하는 이퀄라이저로, 결과적으로 2차원의 이미지로부터 2.5차원의 동적 시뮬레이션이 가능하도록 한다. 여기서 말하는 이미지의 콘텐츠는 이미지가 가지고 있는 텍스트, 캐릭터 등이 될 수 있으며, 캐릭터의 얼굴, 몸과 같이 세부적인 부분도 해당된다.

이미지 내에 존재하는 콘텐츠의 형태를 유지하는 2.5차원의 이퀄라이저를 실시간으로 생성하기 위해서는 몇 가지 해결해야하는 문제점들이 있다. 우선 이미지의 형태를 어떻게 유지 시킬 것 인지에 대한 방안이 필요하다. 본 논문에서는 이미지 콘텐츠의 형태를 유지시키기 위해 사용자가 유지시키고자 하는 부분을 관심영역으로 지정하도록 한다. 관심영역지정은 영상처리 분야에서 특징 분석에 주로 사용되는 마스킹 기법을 활용한다. 지정된 관심영역은 실제 이퀄라이저가 만들어지는 영역으로 콘텐츠의 형태를 유지하는데 사용된다. 다음은 2차원의 이미지로부터 2.5차원의 동적인 움직임이 가능하도록 하는 것이다. 2차원 공간은 이미지가 가지는 Degree of Freedom(DOF)의 제약으로 입체적인 움직임을 줄 수 없기 때문에 이미지를 격자에 맵핑시켜 2차원 공간상의 문제를 해결할 수 있는 환경을 만든다. 이를 기반으로 주파수를 높이 맵에 맵핑시킴으로써 2차원의 이미지로부터 입체적인 움직임이 가능하도록 한다. 이러한 과정을 통해 특정 영역의 형태적 특징을 유지시키는 비주얼 시뮬레이션이 가능하다. 이는 그래픽의 패턴에만 집중하는 기존 방식의 이퀄라이저에서 벗어난 새로운 형태의 이미지 이퀄라이저로써 음악과 함께 감상하는 그림 등에 유용하게 활용될 수 있다.

## 2. 논문의 독창성

기존의 사운드 이퀄라이저와 달리 본 논문이 가지는 독창성은 사운드를 시각적으로 표현하는 ‘시각화 방법’에 있다. 기존 방식의 사운드 이퀄라이저는 대부분 사운드의 주파수와 같은 수치 정보를 막대 그래프형태로 표현하고 있으며, 이는 그래픽 패턴을 애니메이션하는 단순한 형태의 비주얼 시뮬레이션이다. 본 논문에서 제안하는 방법은 사운드와 이미지를 입력받아 소스로 사용하고, 이미지 내의 콘텐츠 형태를 유지하며 이퀄라이저를 생성하는 비주얼 시뮬레이션 형태이다. 이미지 내의 콘텐츠는 사용자가 관심영역으로 지정하며 그림 3에서와 같이 캐릭터의 배경 부분을 관심영역으로 지정하게 되면 배경 부분에만 비주얼 이펙트가 적용되어 맨 오른쪽 그림과 같은 시뮬레이션 결과를 얻게 된다. 기존의 이퀄라이저가 사운드의 특징을 막대 그래프형태로 표현한 사운드 이퀄라이저라면, 본 논문은 사운드의 특징을 분석하여 이미지에 시뮬레이션한 이미지 이퀄라이저라고 할 수 있다.



그림 3. 본 논문에서 제안하는 사운드의 비주얼 시뮬레이션

이를 위해 본 논문에서는 사운드와 이미지를 분석하여 매칭하는 방법을 제안하였고, 2차원의 입력 이미지를 2.5차원의 시뮬레이션으로 표현하기 위해 사운드의 주파수를 높이 맵에 맵핑하는 방법을 제시하였다. 본 시스템에 대한 자세한 설명은 3장 시스템 구성에서 설명한다.

### 3. 논문 구성

본 논문의 구성은 다음과 같다. 2장에서는 본 논문의 관련 연구로 사운드 시각화에 관한 연구와 시각화 작품에 대해서 살펴본다. 3장에서는 본 논문의 시스템 구성을 설명하고, 4장에서는 사운드의 실시간 비주얼 시뮬레이션 방법에 대해 설명한다. 5장에서는 실험 결과에 대해 기술하고, 6장에서는 결론 및 향후 연구 방향에 관하여 논의한다.

## II. 관련연구

우리는 일반적으로 귀를 통해 음악을 들으며 즐거움을 느끼고, 눈을 통해 그림을 보고 감상한다. 하지만 디지털 기술의 발전과 함께 눈으로 듣는 듯한, 귀로 보는 듯한 감정을 느끼려는 다양한 시도들이 활발히 연구되고 있다[1]. 본 장에서는 디지털 기술을 이용해 사운드를 시각적으로 표현하는 연구에 대해 알아보고, 사운드 시각화를 활용한 작품을 살펴본다.

### 1. 사운드 시각화

시각과 청각을 연관 지으려는 시도는 현대에 들어서 새롭게 연구되는 분야가 아니며 디지털 기술이 존재하기 이전부터 고대 철학자, 사상가 및 예술가들에 의해 연구되어 왔다. 고대 물리학자인 뉴턴[2]은 빛을 7가지 색으로 구분하고 각 빛을 음의 파장과 비교하여 음과 색의 유사성을 과학적으로 증명하였다. 그의 연구는 컬러와 사운드가 파장 길이와 주파수로 설명될 수 있다는 개념의 기초가 되었다. 칸딘스키[3]는 음과 색의 연관성에 대해 본격적으로 연구한 화가로서 음악에서 음의 높낮이가 다르고, 악기마다 독특한 음색이 있듯이 색채도 저마다 느낌이 다르다는 것을 밝혀냈다. 또한, 각각의 악기는 저마다 가지고 있는 색깔이 있다고 하며 트럼펫은 노란색, 비올라는 오렌지색, 첼로는 파란색, 바순은 보라색, 오르간은 초록색, 튜바나 북은 빨간색의 느낌을 유발한다고 하였다.

국내에서도 음과 색의 연관성에 관한 연구로 소리와 색을 1:1로 대응시켜 소리를 색채로 표현하려는 연구들이 이루어지고 있다[4][5][6]. 이 연구에서는 색채와 소리는 물리적으로 공진, 증폭, 간섭, 상쇄 등의 파동의 속성을 공유하고 있고, 도, 미, 솔의 주파수 비율이 빨강, 초록, 파랑의 주파수 비율과 거의 일치한다는 원리를 이용해 소리와 색채를 연결하였다(그림4).

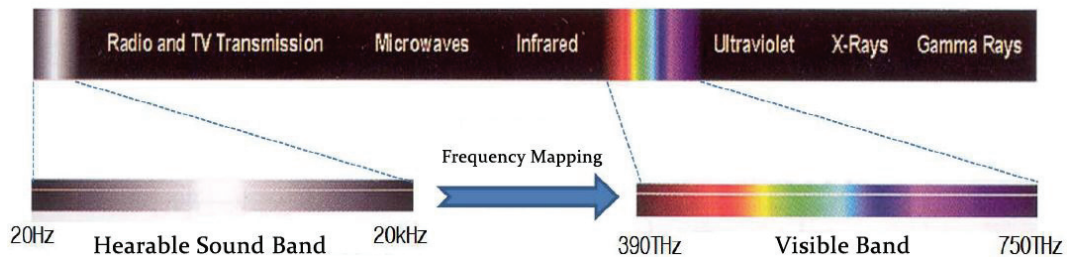


그림 4. 가청 주파수와 가시 주파수 매칭을 통한 음-색 대응

가청 주파수와 가시 주파수의 매칭 방법을 기반으로 Kim[7]는 인간의 공감각적 기능을 모방하는 지능로봇의 개발을 위한 첫 번째 단계로 색-음 상호 변환 시스템을 구축하였다. 그림 5에서와 같이 RGB의 컬러 영상을 HSI 영상으로 변환 분리하고 각각의 색상, 채도 및 명도 값을 음의 3요소인 음의 높이, 음계 및 옥타브에 각각 맵핑시켜 영상으로부터 음원을 생성하였다.

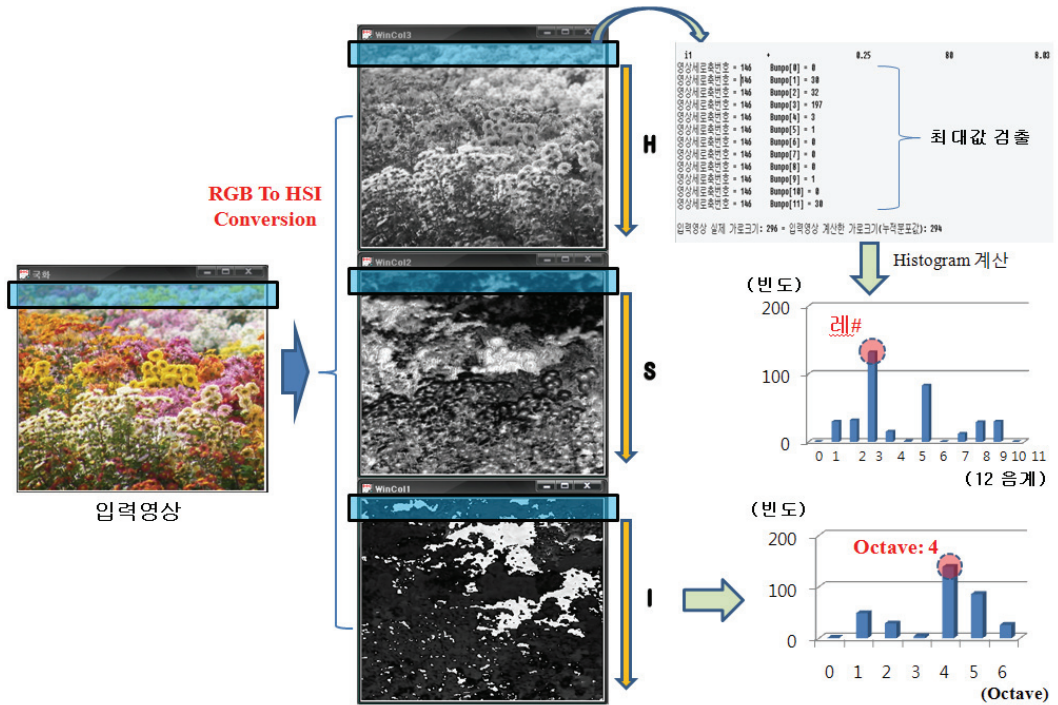


그림 5. HSI 컬러 히스토그램을 이용한 색-음 변환

Kim[8]은 사운드 데이터를 사용하여 게임 플레이상에 지형을 생성하는 방법에 대해 연구하였다. 입력받은 사운드를 푸리에 변환하여 취득한 데이터를 분석 및 재가공하여 시간에 따라 변화하는 3차원 지형을 생성하였다(그림6). 이는 화면과 입출력되는 사운드 간의 동기화를 통해 사용자들에게 공감각적인 상호작용을 제공하였다.



그림 6. 사운드에 의해 생성된 지형 (방설, 사막, 바다)

Kim[9]등은 SOUND BOUND라는 사용자 인터랙티브한 그래픽 이퀄라이저를 제안하였다. 제안하는 이퀄라이저는 기존의 그래픽 이퀄라이저와 유사하게 주파수를 그래프로 보여주고, 뿐만 아니라 사용자가 터치 인터랙션으로 아이콘을 생성하고 아이콘들이 주파수 그래프에 닿으면서 음향 효과를 생성할 수 있도록 하였다(그림7).

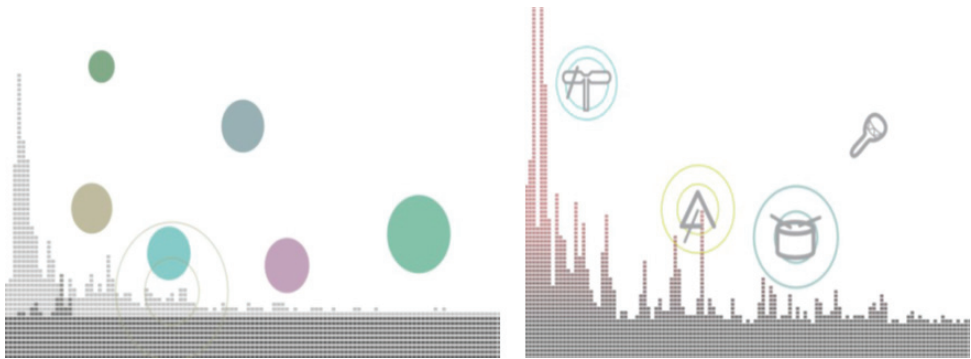


그림 7. Sound Bound 결과

Seo[10]등은 이미지를 사용한 사운드 그래픽 이퀄라이저의 구성에 대해 연구하였다. 기존 방식의 막대그래프의 형태에서 벗어나 새로운 시각적 형태를 지닌 사운드 그래픽 이퀄라이저로써 사용자 친화적인 GUI 형태의 구축 방법을 제시하였다. 그림8은 입력 이미지로 장미꽃 이미지를 사용한 결과 영상으로 음량의 크기에 따라 이미지의 개수와 크기를 증.감하고 주파수 대역에 따라 이미지의 종류를 달리하여 전체적인 이미지에서 음원의 개별적 정보를 알 수 있도록 하였다. 하지만 기존방식의 이퀄라이저와 같이 여전히 그래픽의 패턴에 초점을 맞추고 있다.



그림 8. 사운드 그래픽 이퀄라이저 결과

Kim[11]은 사운드 중 한국의 전통악기인 가야금에 초점을 맞추어 새로운 가야금 연주 인터페이스와 가야금 소리를 시각화 하였다. 오랜 시간의 연습을 통해 익힌 숙련된 기술로 연주해야하는 가야금을 그림9의 왼쪽과 같이 쉽고 새로운 인터페이스로 변형하여 연주를 가능하게하고 소리와 붓 획을 매치시켜 한국화 느낌의 결과를 생성하였다.



그림 9. 가야금 연주 인터페이스(좌)와 시각화 결과(우)

Yoo[12]는 지능형 음악 분수 제작 시스템을 제안하였다. 기존에 생성되었던 음악분수의 시나리오를 분석하여 데이터베이스화 시킨 후, 이를 새롭게 합성하여 시나리오를 생성한다. 또한 음악분수 시나리오 생성에서 중요하게 요구되는 음악과 물 모양의 동기화를 위하여 임의의 음악이 들어왔을 때, 음악의 비트를 추출한 후 비트 단위로 시나리오를 합성하게 된다. 그림 10은 음악분수 시물레이션 결과를 보여준다.

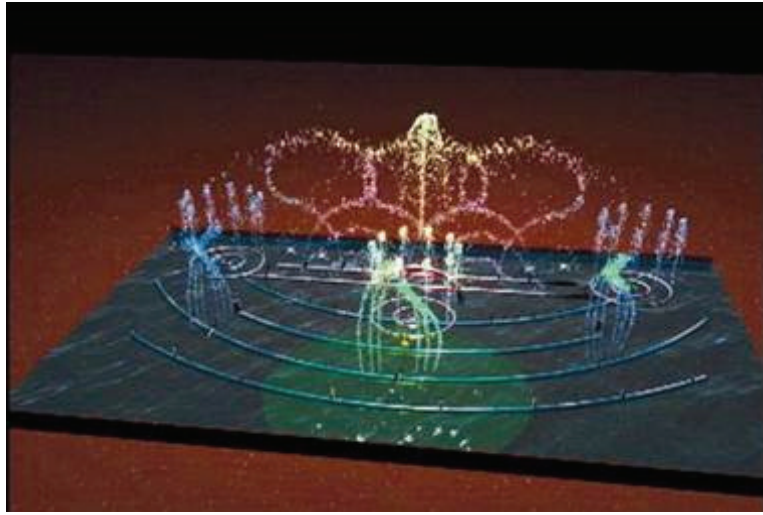


그림 10. Creating Musical Fountain Shows 결과

Kang[13]등은 사용자가 그린 그림을 음악으로 만들어주는 인터랙티브 인터페이스를 제안하였다. 사용자가 그린 그림이 센서박스를 통과하는 동안 그림의 잉크 방향과 두께 등을 인식하고 이를 음악의 특징(음계, 볼륨, 템포)으로 변환하여 음악을 만들어내고 동시에 그림이 그려져 있는 부분이 센서 위를 지날 때 색조명이 켜져 시각적인 효과도 생성한다(그림11).

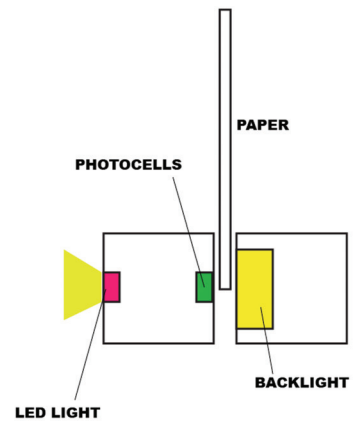
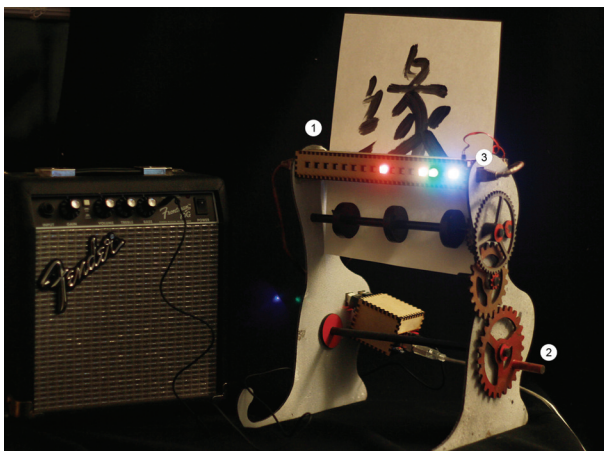


그림 11. Harmonic Paper 인터페이스와 구조

## 2 오디오 시각화 작품 분석

### 2.1 Piano: as image media

이 작품은 실제 피아노 연주와 함께 3차원의 이미지들이 디스플레이되어 악기 소리를 이미지화하는 작품이다(그림12). 사용자가 트랙볼을 움직이면 피아노 아래쪽에 프로젝션되는 그리드 위에 트랙볼이 움직이는 위치에 따라 점이 찍히게 된다. 이 점들은 피아노 쪽으로 이동하여 피아노 건반에 부딪히면서 피아노 소리를 내는 것과 동시에 다양한 색상과 형태의 이미지가 디스플레이 된다.

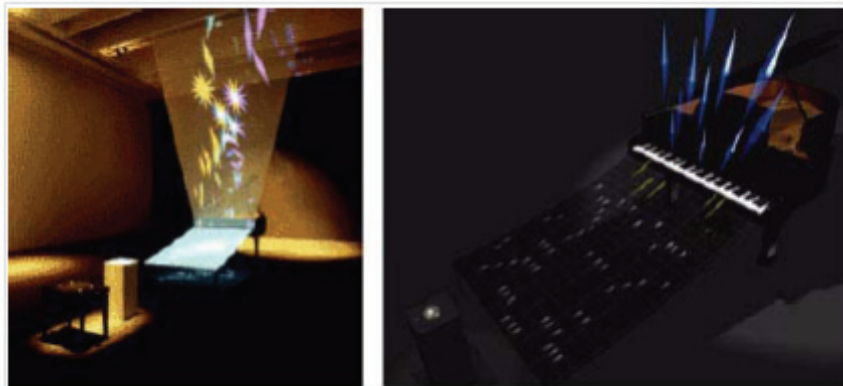


그림 12. <Piano: as image media> 1995. Toshio Iwai

### 2.2 Rhyme Modulation - nong(弄)

실시간 미디어 퍼포먼스 작품 ‘Rhyme Modulation - nong, 2006, 오창근’은 그림 13에서 보이는 것과 같이 실제 가야금 연주 영상을 카메라로 촬영하고 연주에 맞춰 실시간으로 영상을 변화시키는 작품이다. 가야금 즉흥 연주를 마이크를 통해 디지털화하고 그 데이터를 실시간 영상 변조에 이용하며, 연주자를 촬영하는 카메라 영상을 작품의 소스영상으로 사용했다.



그림 13. Rhyme Modulation - nong(弄) 2006. 오창근

### 1.2.3 김수정, “공명의 Deep Sea”

우리나라 전통음악을 바탕으로 새로운 아이디어를 결합한 공연을 하는 퓨전 타악기 그룹 ‘공명’의 공연으로, 실시간으로 연주되는 아날로그 신호를 분석하여 시각화한다. 네 명의 연주자가 만들어내는 화음을 영상으로 해석해서 관객에게 보여줌으로써 공연의 몰입도를 높였으며 곡의 흐름에 따라 영상을 분할하여 상영함으로써 곡의 묘미를 더하였다. 그림 14는 ‘공명’의 공연 모습을 보여준다.



그림 14. 공명의 Deep Sea, 2008

#### 2.4 김윤태, “허클베리핀의 옐로우 콘서트”

두 명의 드럼 연주자가 즉흥적으로 드럼을 연주하면 드럼의 비트와 음량에 반응하여 자동으로 이미지가 변화되는 작품이다(그림15). 이미지는 드럼 연주뿐 만 아니라 관객의 박수 소리가 드럼 음량에 달할 정도로 클 경우에도 변화되는데, 이를 통해 관객의 호응을 더욱 이끌어 낼 수 있으며 음악을 듣는 청중에게 소리뿐 만 아니라 시각적인 효과도 크게 영향을 미친다는 것을 보여준다.



그림 15. 허클베리핀의 옐로우 콘서트, 2007

#### 2.5 The Sancho Plan, “Spacequatica” (2008)

세 명의 연주자가 전자 드럼 패드로 연주를 하면 그 소리에 맞추어 스크린의 3D 오브젝트가 변형되는 작품이다. 연주에 의해 영상이 변화됨으로써 연주자와 영상간의 상호 소통이 가능한 의미 있는 퍼포먼스가 이루어졌고 소리에 반응하는 오브젝트가 살아 움직이는 듯한 감성적인 느낌을 전달함으로써 독특한 영상에 관객들이 매료되었다. 이를 통해 음악을 들을 때 청각적인 요소뿐 만 아니라 시각적인 요소도 중요한 역할을 한다는 것을 알 수 있다. 그림 16은 실제 공연모습으로 연주자가 드럼을 두드려 소리를 내면 왼쪽아래에 있는 오브젝트가 오른쪽아래의 오브젝트 모형으로 바뀌게 된다.

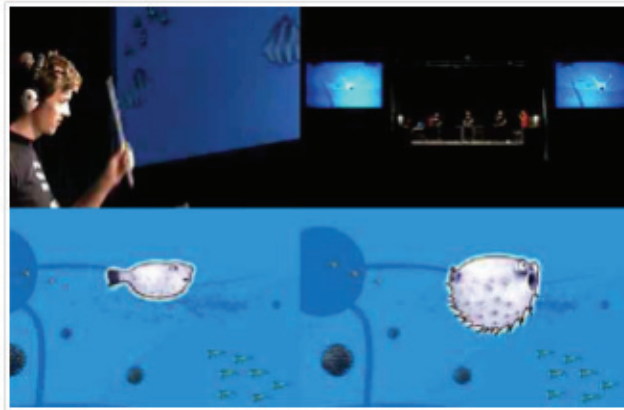


그림 16. Spacequatica 오스트리아 공연, 2008

## 2.6 Picturedisko

Picturedisko는 이미지의 픽셀정보를 미디신호로 변화시켜 소리로 표현하는 작품이다. 이미지는 온라인 사진 공유 커뮤니티 사이트인 flickr.com에서 실시간으로 이미지들을 가져와서 사용하며, 그림 17의 오른쪽에서 이미지를 선택하면 왼쪽에 있는 턴테이블위에 이미지가 올려져 이미지의 픽셀정보에 따라 다양한 소리를 내며 연주되는 작품이다.



그림 17. Picturedisko 2006. Mario Klingemann

### Ⅲ. 시스템 구성

본 논문은 이미지 내에 있는 특정 영역의 형태적 특성을 유지하는 2.5차원 동적 시뮬레이션을 위한 프레임워크를 제안한다. 제안하는 시스템은 크게 이미지와 사운드를 분석하는 단계와 시뮬레이션 단계로 나뉘어진다.

이미지와 사운드를 분석하는 단계에서는 입력 이미지의 형태를 유지하기 위해 사용자가 관심영역을 지정하며, 관심영역을 지정하기 위한 방법으로 마스킹 기법이 사용된다. 동적인 움직임을 위해서는 입력 사운드의 주파수 특징을 추출하기 위해 고속 푸리에 변환을 사용한다. 시뮬레이션 단계는 다시 세부적으로 이미지의 격자 맵핑, 높이 맵 기반 주파수 맵핑, 그리고 컨볼루션 과정으로 나뉘어진다. 그림 18은 본 논문의 시스템 개요를 보여주며, 본 장에서는 시스템 개요의 세부적인 과정을 설명한다.

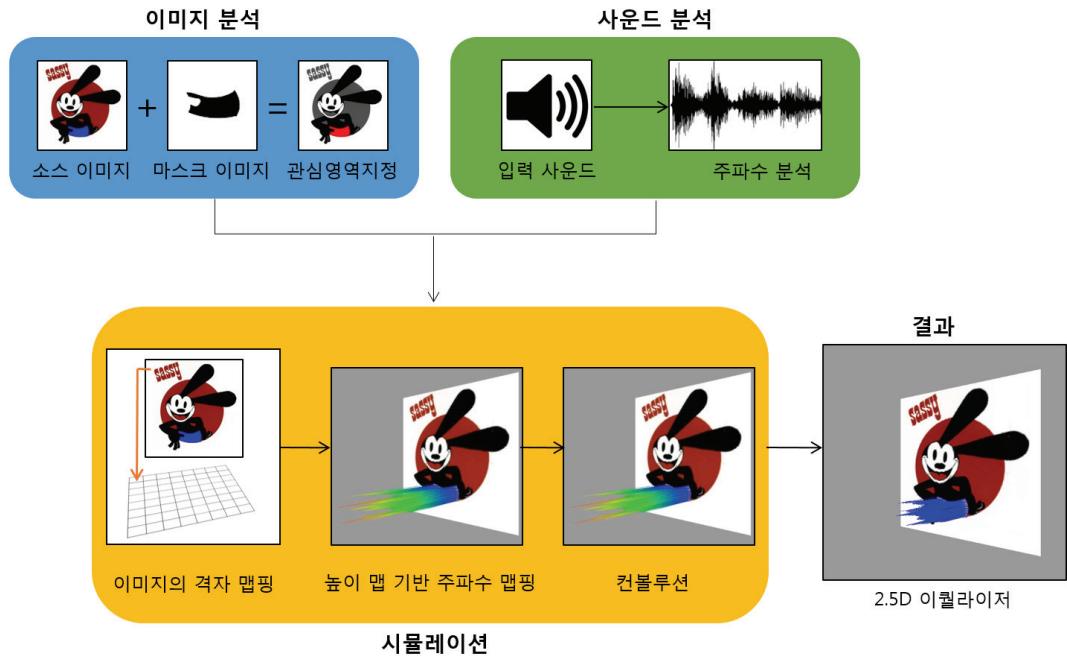


그림 18. 시스템 개요

## 1. 이미지와 사운드 분석 단계

이미지와 사운드 분석 단계는 이미지의 관심영역 지정과 사운드 주파수 분석 과정으로 나누어진다. 본 논문에서는 이미지 내 특정 영역의 형태를 유지시키기 위한 방법으로 마스킹 기법을 사용한다. 마스킹은 영상 처리 분야에서 특징을 추출하거나 형태를 분석하기 위해 사용되는 방법으로 주로 이진 영상을 사용한다. 소스 이미지와 마스크 이미지의 논리 연산을 통해 관심영역을 지정하며 이를 통해 이미지의 형태적 특징을 유지시킬 수 있다. 동적인 시물레이션을 위해서는 사운드 이퀄라이저를 이용하는데 이퀄라이저를 생성하기 위해서는 사운드의 분석이 우선시 되어야한다. 입력되는 사운드의 주파수 성분을 분석하기 위해서 고속 푸리에 변환을 사용한다. 고속

푸리에 변환을 거치면 사운드의 주파수 성분과 진폭 값을 추출할 수 있으며 이는 시뮬레이션 단계에서 이퀄라이저를 생성할 때 사용된다. 이미지와 사운드를 분석하는 방법은 4장에서 자세하게 설명한다.

## 2. 시뮬레이션 단계

시뮬레이션 단계는 이미지와 사운드 분석 단계에서 추출한 데이터를 기반으로 실시간 동적 시뮬레이션이 이루어지는 단계로 이미지의 격자 맵핑, 높이 맵 기반의 주파수 맵핑, 그리고 컨볼루션의 세부 과정으로 나누어진다.

### 이미지의 격자 맵핑

입력 영상은 2차원의 이미지로 2차원 공간은 DOF의 제약을 가지고 있어 움직임을 표현해야하는 동적 시뮬레이션에는 적합하지 않다. 이러한 공간상의 문제를 해결하기 위해 격자 맵핑 과정을 수행한다. 입력 이미지의 픽셀은 격자와 1:1로 맵핑되어 2차원이라는 도메인의 제약을 해결한다.

### 높이 맵 기반 주파수 맵핑

사운드 주파수 분석 후 추출된 데이터를 기반으로 이퀄라이저를 생성한다. 이퀄라이저는 사운드의 주파수 성분에 의해 생성되는데 추출된 주파수 데이터는 보통 관심영역의 픽셀 수 보다 적어(본 논문에서는 512크기의 데이터로 추출) 주파수 값과 관심영역 상의 픽셀을 1:1로 매칭 시킬 수 없다. 제한된 수의 주파수 데이터를 관심영역에 표현하기 위해 모듈러 방식을 사용한다. 결과적으로 x축과 y축만 가지는 2차원의 이미지에서 2.5차원의 동적 시뮬레이션이 가능하도록 한다.

## 컨볼루션

제한된 수의 주파수 데이터를 관심영역의 픽셀 데이터에 맵핑시키면 일정한 패턴 현상이 일어나게 된다. 음악에서 특정 주파수 영역이 높게 나타나는 경우에는 반복되는 패턴으로 인해 부자연스러운 시뮬레이션을 생성할 가능성이 있다. 이를 해결하기 위해 가우시안 필터를 사용하여 컨볼루션 시키는 과정을 거친다.

## IV. 사운드의 실시간 비주얼 시뮬레이션

### 1. 이미지와 사운드 분석

본 단계에서는 입력 이미지의 관심영역을 지정하는 방법과 사운드의 주파수 특징을 실시간으로 분석하는 방법에 대해 설명한다. 관심영역을 지정하기 위한 방법으로 마스킹 기법을 사용하여 이미지에서 유지시키고자 하는 영역이 지정된다. 사운드의 주파수 특징 분석에서는 기본적인 음향 신호와 주파수에 대해 설명한 뒤, FFT 알고리즘을 사용하여 입력 사운드의 주파수 성분을 분석하는 방법에 대해 알아본다. 주파수 분석 후 추출된 데이터는 시뮬레이션 과정에서 이퀄라이저를 생성하는데 사용된다.

#### 1.1. 이미지 관심영역 지정

기존의 이퀄라이저는 기본적으로 막대그래프의 형태로 나타내며 각 막대는 주파수와 데시벨을 나타내는 형식을 지닌다. 현재 재생중인 사운드의 주파수와 크기를 그래프 형식으로 보여주는 기존의 이퀄라이저의 방식에서 벗어나 사용자의 피드백을 받아 기존의 방식과 차별된 이퀄라이저를 생성하기 위해 관심영역을 지정한다. 입력 이미지에서 사용자가 유지시키고자하는 특정 영역을 관심영역으로 지정하기 위해 마스킹 기법을 이용한다. 마스킹 기법은 영상처리 분야에서 이미지의 특징 추출, 영역 분석 등에 사용되는 기법으로, 마스킹 처리를 위한 마스크 이미지는 주로 이진 영상을 사용한다.



그림 19. 마스킹 과정

그림 19는 마스킹 과정을 보여준다. 소스 이미지와 유지시키고자 하는 영역을 나타낸 이진 영상과의 논리 연산을 통해 관심영역을 지정할 수 있다. 관심영역으로 지정된 부분은 이퀄라이저 생성 시 사용되어 사용자가 지정한 관심영역의 콘텐츠를 유지시킬 수 있다. 그림 20은 마스킹을 통한 관심영역 선택 과정을 보여준다. 그림 20의 마스크 이미지는 이미지 내의 콘텐츠를 나누어 생성한 이진 영상이며, 원하는 영역을 선택하여 사용자가 원하는 관심영역을 설정할 수 있다. 하나의 마스크 이미지는 하나의 관심영역으로 지정될 뿐만 아니라, 다수의 마스크 이미지를 결합하여 새로운 형태의 관심영역을 지정할 수 있다. 예를 들어 귀와 얼굴을 나타내는 마스크 이미지는 각각 관심영역으로 지정될 수 있지만, 귀와 얼굴 마스크 이미지를 결합하여 붉은 네모로 표시된 것과 같은 새로운 형태의 관심 영역을 지정할 수 있다. 이는 사용자마다 원하는 관심영역이 다르기 때문에 사용자 맞춤형 관심영역을 지정하기 위함이다. 위 과정을 통해 사용자의 피드백을 받아 이퀄라이저를 생성할 수 있다는 점에서 기존 이퀄라이저와 차별되는 점이다.

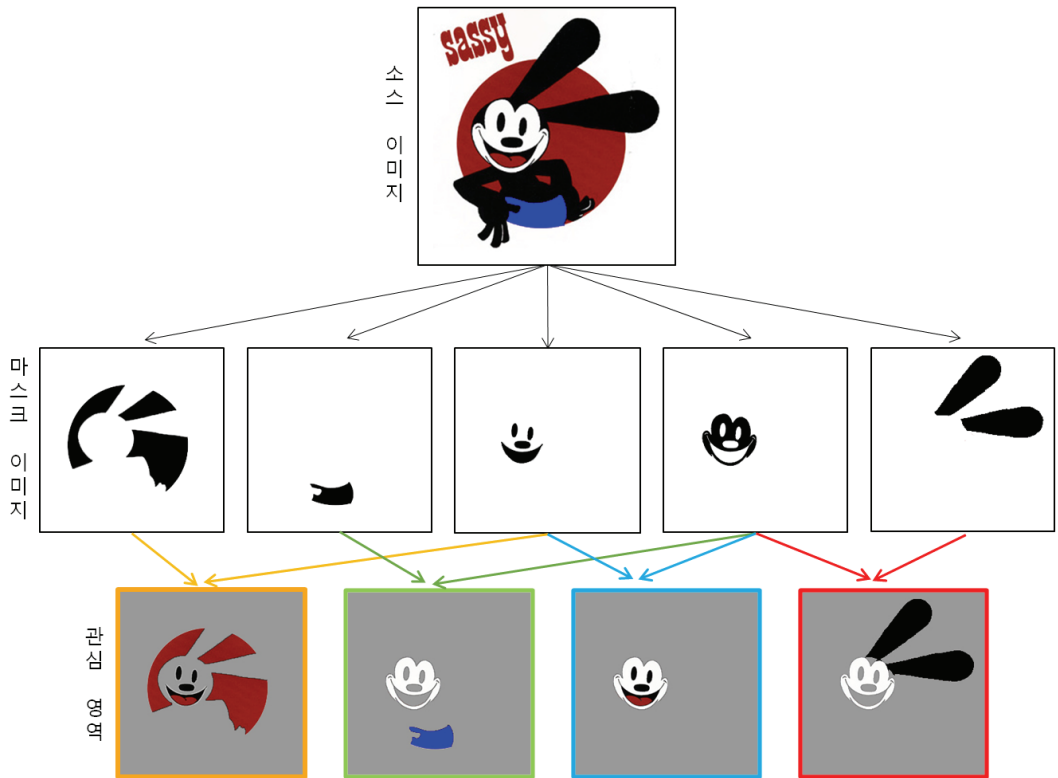


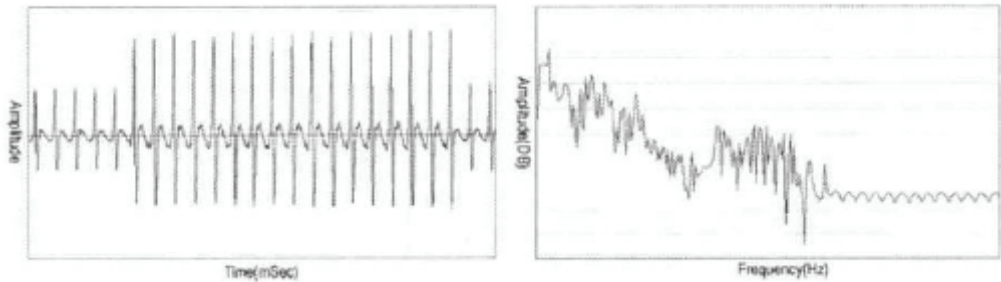
그림 20. 마스크를 통한 관심영역 선택 과정

## 1.2. 사운드 주파수 분석

### 1.2.1 음향 신호의 분석

음향 신호는 크게 시간 도메인에서 표현하는 방법과 주파수 도메인에서 표현하는 방법으로 나눌 수 있다. 시간 도메인은 시간 변화에 따른 신호의 크기(진폭) 변화를 나타내는 방식으로 그림 21의 (a)와 같이 가로축이 시간을 나타내고 세로축은 진폭(크기)을 나타낸다. 주파수 도메인은 주파수에 대한 최대 진폭을 나타내는 방식으로 진폭과 주파수의 관계를 표현한다. 주파수 도메인은 그림 21의 (b)와 같이 가로축이 주파수를 나타내며, 세로축이 진폭을 나타낸다. 모든 신호는 시간 도메인과 주파수 도메인으로 분석하

여 나타낼 수 있으며, 분석된 데이터를 사용하여 다양한 분야에 응용될 수 있다.



(a) 시간 도메인

(b) 주파수 도메인

그림 21. 시간 도메인과 주파수 도메인에서의 음향 신호

주파수란 주기적인 현상이 단위시간 동안 몇 번 일어났는지를 나타내는 것으로 주파수 단위로 헤르츠(Hz)를 사용한다. 즉, 1Hz는 1초에 1주기가 나타나는 것이고, 2Hz는 1초에 2주기가 나타나는 것이다(그림 22).

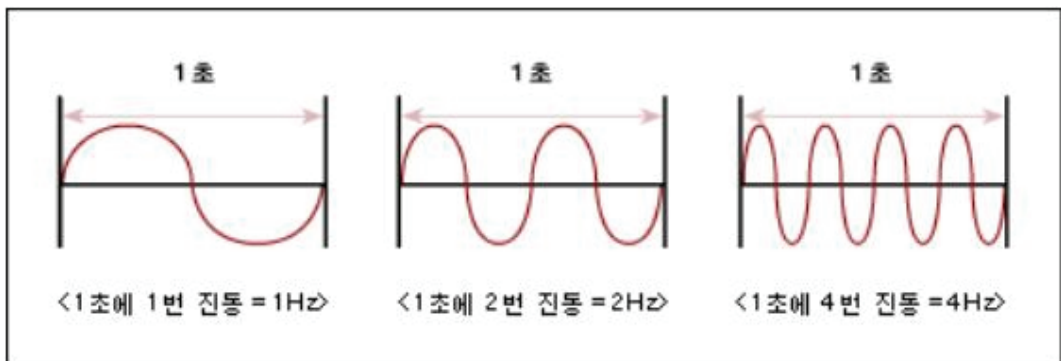


그림 22. 주파수의 주기

음향 신호를 분석했을 때 주파수의 파형에 따라 소리의 크기(데시벨)와 음

의 높낮이는 분석할 수 있다. 일반적으로 주파수의 진폭이 크면 큰 소리가 나며 진폭이 작으면 작은 소리가 난다. 또한 주파수의 파장에 따라 음의 높낮이가 달라지는데, 파장이 크면 저주파 대역의 낮은 소리가 나며, 파장이 작으면 고주파 대역의 높은 소리가 나게 된다(그림 23).

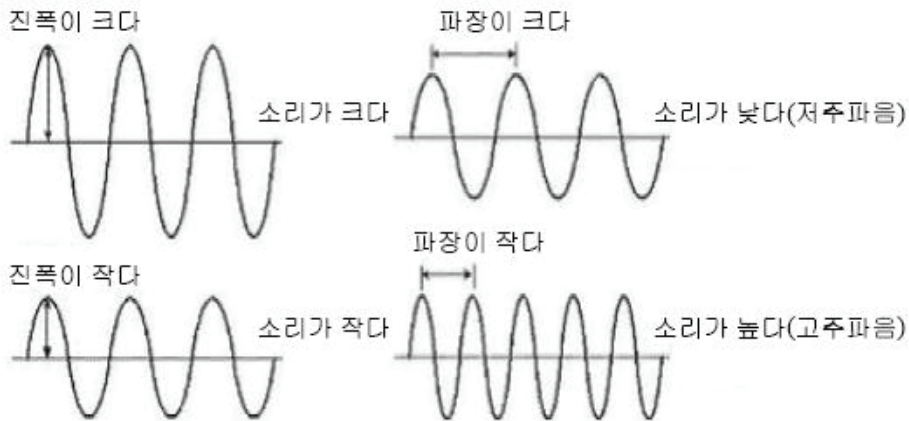
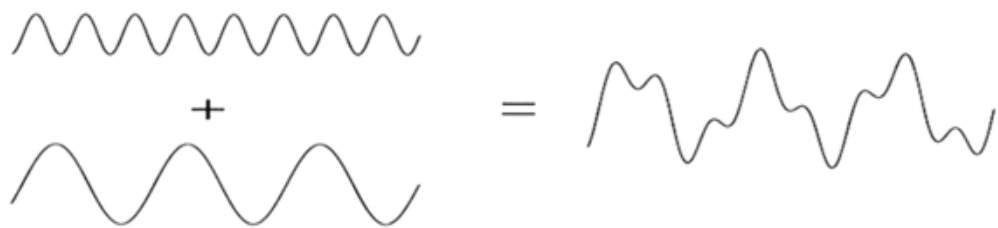


그림 23. 주파수 파형에 따른 소리의 특성

주파수는 어떠한 파형도 정현파의 합으로 표현될 수 있다. 즉, 하나의 음을 표현하는 정현파들이 모여 주파수가 되며, 반대로 주파수를 분석하면 여러 개의 정현파들로 구성되어 있다는 것을 알 수 있다(그림 24).



(a) 각각의 정현파

(b) 합쳐진 신호

그림 24. 주파수 표현

이러한 주파수는 푸리에 변환(Fourier Transform)을 통해 분석할 수 있으며 푸리에 변환에 대해서는 다음 단계에서 자세히 설명한다.

### 1.2.2. 푸리에 변환(Fourier Transform : FT)

푸리에 변환은 시간 도메인 상의 신호를 주파수 도메인의 신호로 변환하는 것이다. 모든 파형은 단순한 정형파의 합으로 표현할 수 있다는 이론에 기반 한 것으로 시간 도메인의 신호에 코사인(cosine)과 사인(sine)을 곱하여 주파수별로 분해하는 것을 말한다(그림 25). 하지만 이러한 푸리에 변환은 디지털 음향 신호를 분석하는데 적합하지 않기 때문에 디지털 신호와 같이 이산적인 수열에 적용되는 푸리에 변환을 사용하여야 한다. 이를 이산 푸리에 변환(DFT)라고 하며 그 식은 식(1)과 같다.

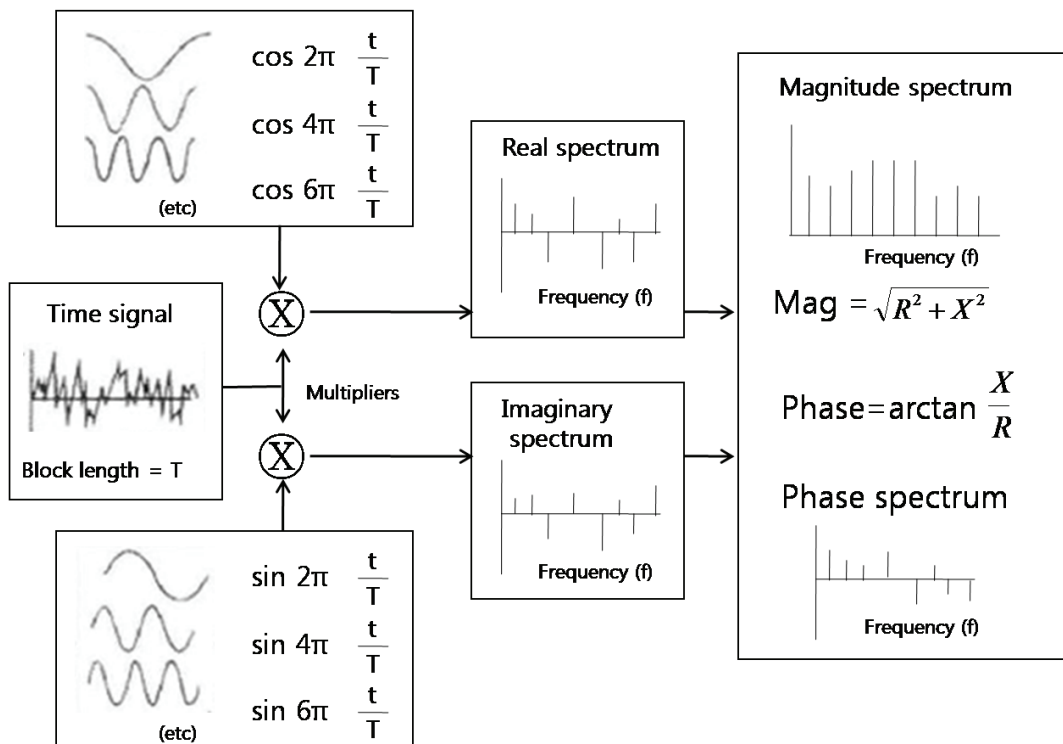


그림 25. 푸리에 변환

$$G_{\sigma}(x, y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{x^2 + y^2}{2\sigma^2}\right] \quad (1)$$

이산 푸리에 변환은 많은 연산량을 필요로 하기 때문에 음향 신호를 실시간으로 분석하는데 적합하지 않다. 이러한 단점을 보완하기 위해서 등장한 것이 고속 푸리에 변환이다. 고속 푸리에 변환은 반복되는 연산을 줄여 이산 푸리에 변환을 고속으로 수행하는 알고리즘으로 고속 푸리에 변환을 사용하여 주파수 분석하는 방법을 FFT 분석이라고 한다. 고속 푸리에 변환을 통해 분석된 결과는 이산 푸리에 변환의 결과와 동일하다.

본 논문에서는 FFT를 사용하여 입력 사운드의 주파수 성분을 분석한다. FFT를 한 후 얻어지는 주파수 성분과 각 주파수에 따른 진폭으로 이퀄라이저를 생성한다. 그림 26은 입력 사운드에 FFT를 사용하여 주파수를 분석한 후 스펙트럼으로 표현한 것이다.

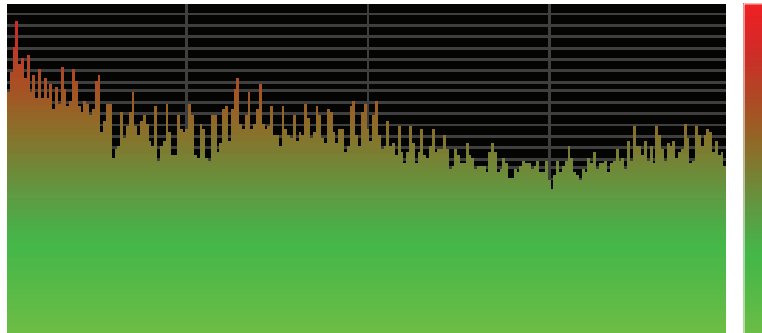


그림 26. 입력 사운드의 스펙트럼

## 2. 격자 기반 시뮬레이션

본 장에서는 관심영역과 추출된 사운드 주파수 특징을 이용한 시뮬레이션 방법에 대해 설명한다. 입력 이미지가 가지는 공간상의 제약을 해결하여 2차원공간에서의 입체적인 움직임을 위해 이미지를 격자에 1:1로 맵핑하고 이를 기반으로 이퀄라이저를 생성하여 동적 시뮬레이션이 가능하도록 한다. 마지막으로 위해 컨볼루션 단계를 거쳐 입력 사운드와 연동되는 2.5D 동적 시뮬레이션이 가능하다.

### 2.1. 이미지의 격자 맵핑

입력 영상은 x축과 y축, 두 개의 축만을 가지는 2차원 공간의 이미지이다. 2차원 공간은 Degree of Freedom(DOF)의 제약으로 입체적인 움직임을 표현해야 하는 동적인 시뮬레이션에 적합하지 않다. 그림 27은 2차원 공간상의 이미지와 DOF를 보여준다.

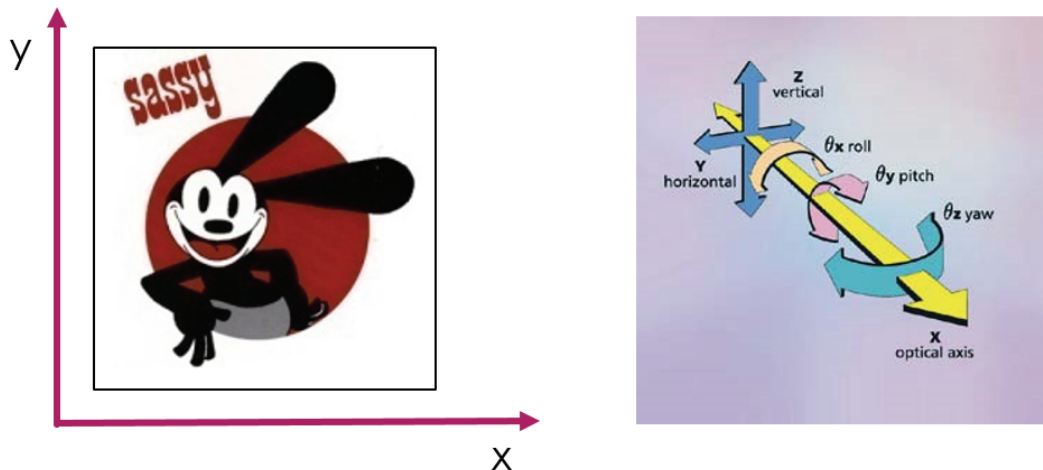


그림 27. 2차원 공간상의 이미지(좌) Degree of Freedom(우)

2차원 공간에서 입체적인 움직임을 표현하기 위해 입력 이미지를 격자에 맵핑시킨다. 그림 28과 같이 이미지의 픽셀과 격자가 1:1로 맵핑되어 각 픽셀단위로 움직임을 제어할 수 있다. 이를 통해 2차원 공간이 가지는 제약을 해결하여 동적 시뮬레이션이 가능하도록 한다. 입력되는 소스 이미지는 그림 28과 같이 격자에 1:1로 맵핑된다.

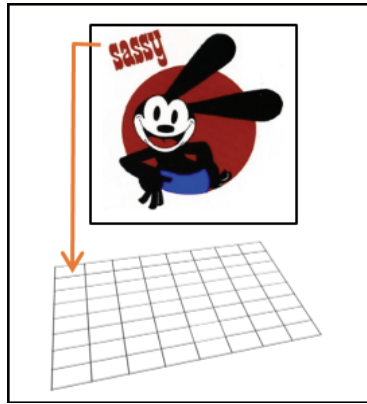


그림 28. 이미지의 격자 맵핑

## 2.2. 높이 맵 기반 주파수 맵핑

현재 재생중인 사운드를 고속 푸리에 변환하여 추출한 주파수 데이터를 실시간으로 관심영역의 픽셀에 맵핑시킨다. 추출된 주파수 데이터는 제한된 크기의 데이터를 가지는 반면, 관심영역의 픽셀 데이터는 사용자가 지정하는 관심영역에 따라 그 크기가 유동적이다(그림29). 따라서 그림 30에서 보이는 바와 같이 주파수 데이터는 관심영역의 픽셀 수와 1:1 매칭 되지 않는다.

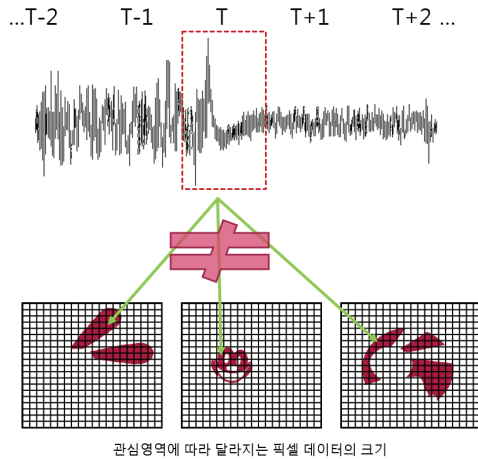


그림 29. 주파수 데이터와 관심영역 데이터의 불일치

이를 해결하기 위해 주파수 데이터를 관심영역 내 픽셀에 맵핑시킨다. 하지만 픽셀 데이터는 관심영역으로 지정되는 부분에 따라 기하급수적으로 커지기 때문에 주파수 데이터의 크기보다 훨씬 많은 데이터를 갖는다. 따라서 주파수 데이터를 관심영역의 픽셀 데이터 전체에 분포하는 방법이 필요하다.

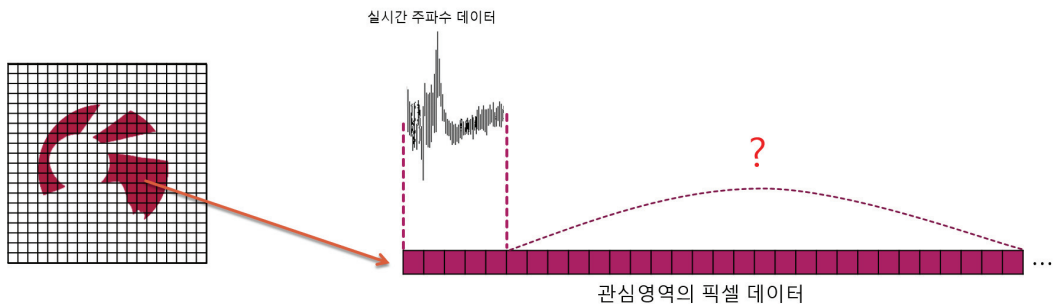


그림 30. 주파수 데이터와 관심영역 내 픽셀 맵핑의 문제점

위 문제를 해결하기 위해 본 논문에서는 모듈러 방식을 이용한다. 모듈러

방식은 패턴 인식 분야에서 데이터의 패턴 분석을 위한 방법으로 어떠한 값도 일정한 패턴으로 분석할 수 있다는 장점을 가진다. 그림 31과 같이 관심 영역의 픽셀 데이터와 주파수 데이터의 모듈러 연산을 통해 주파수 데이터를 관심영역 전역에 분포시킨다.

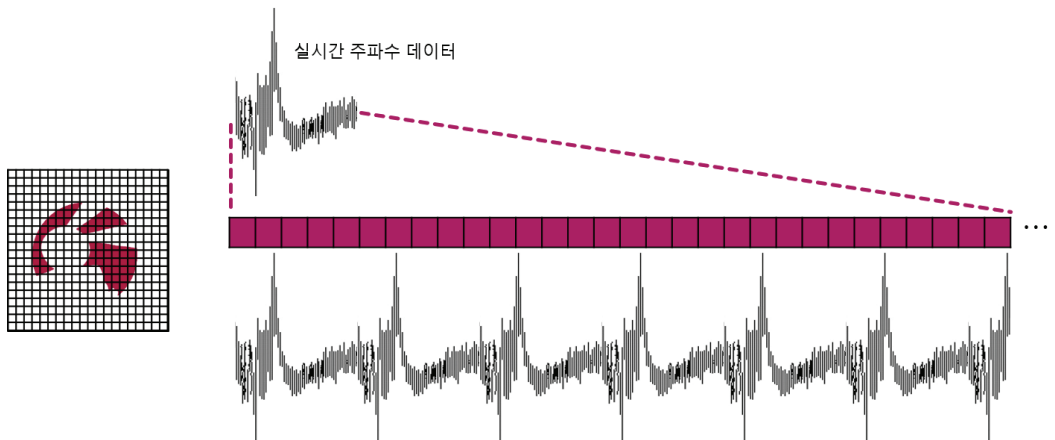


그림 31. 모듈러 방식을 이용한 주파수 데이터 - 관심영역 맵핑

위 과정을 통해 입력 이미지 전체에 대하여 이퀄라이저를 생성하는 것이 아니라 관심영역에만 주파수 데이터를 효과적으로 맵핑하여 관심영역의 형태적 특징을 유지하는 이퀄라이저를 생성할 수 있으며, 이는 사용자 취향에 따른 다양한 이퀄라이저를 생성할 수 있다는 장점을 가진다. 그림 32는 사용자 관심영역에서의 이퀄라이저 생성을 나타내는 그림이다.

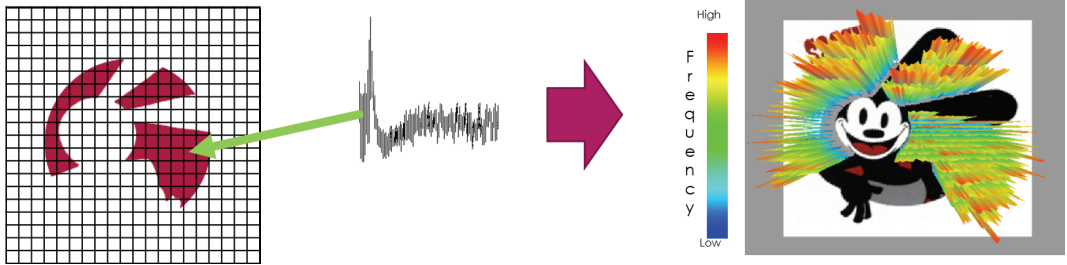


그림 32. 관심영역의 형태적 특징을 유지하는 이퀄라이저 생성

### 2.3. 컨볼루션

모듈러 연산을 통해 픽셀에 맵핑된 주파수 데이터는 그림 33과 같이 일정한 패턴을 가지게 되며 이는 입력 음악에서 특정영역의 주파수 값이 큰 경우에 시뮬레이션동안 부자연스러운 형태를 나타내게 된다.

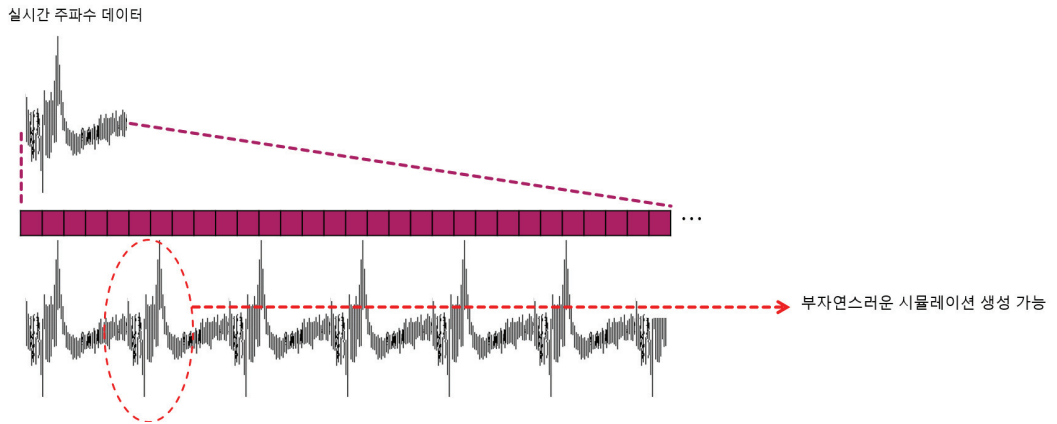


그림 33. 일정한 패턴을 가지는 주파수 데이터의 문제점

이러한 문제점을 해결하기 위해 본 논문에서는 가우시안 필터를 사용하여 컨볼루션하는 단계를 거친다. 그림 34는 가우시안 함수의 그래프를 나타내며 가우시안 필터는 필터내의 평균이 0인 정규 분포를 가지며 그림 35와 같은 마스크 형태를 가진다.

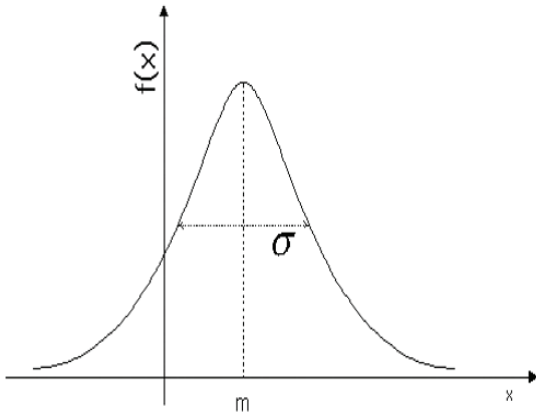


그림 34. 가우시안 함수 그래프

$$\begin{bmatrix} \frac{1}{32} & \frac{1}{32} & \frac{1}{16} & \frac{1}{32} & \frac{1}{32} \\ \frac{1}{32} & \frac{1}{16} & \frac{1}{8} & \frac{1}{16} & \frac{1}{32} \\ \frac{1}{16} & \frac{1}{8} & \frac{1}{4} & \frac{1}{8} & \frac{1}{16} \\ \frac{1}{32} & \frac{1}{16} & \frac{1}{8} & \frac{1}{16} & \frac{1}{32} \\ \frac{1}{32} & \frac{1}{32} & \frac{1}{16} & \frac{1}{32} & \frac{1}{32} \end{bmatrix}$$

그림 35. 5x5 가우시안 마스크

이는 데이터 값을 주변 데이터를 미분한 값으로 대체하여 높이 맵에 가우시안 필터를 적용했을 경우, 높이 맵을 스무딩하여 좀 더 자연스러운 이퀄라이저를 생성할 수 있다.

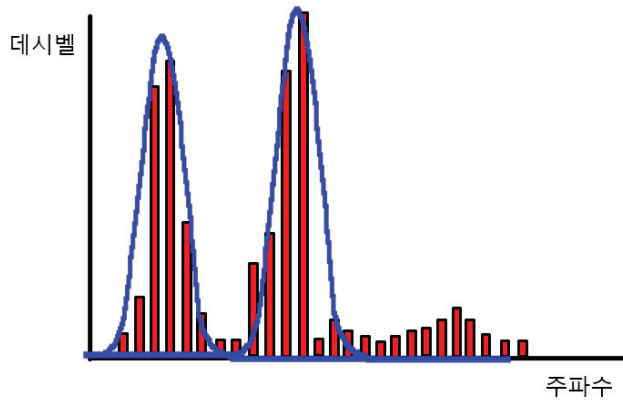


그림 36 가우시안 필터를 사용한 컨볼루션

## V. 실험결과

본 장에서는 제안하는 방법을 사용하여 시뮬레이션 한 결과를 보여준다. 그림 37은 시뮬레이션을 위한 마우스 캐릭터 소스 이미지와 관심영역의 형태를 유지하기 위한 마스크 이미지이다. 시뮬레이션에 사용된 사운드는 여성의 음성 사운드가 사용되었다. 마우스 캐릭터 시뮬레이션에서는 512크기의 주파수 데이터가 10,250개의 관심영역 픽셀에 맵핑되었다. 입력 사운드의 주파수에 따라 달라지는 시뮬레이션 결과는 그림 38과 39에서 알아본다.

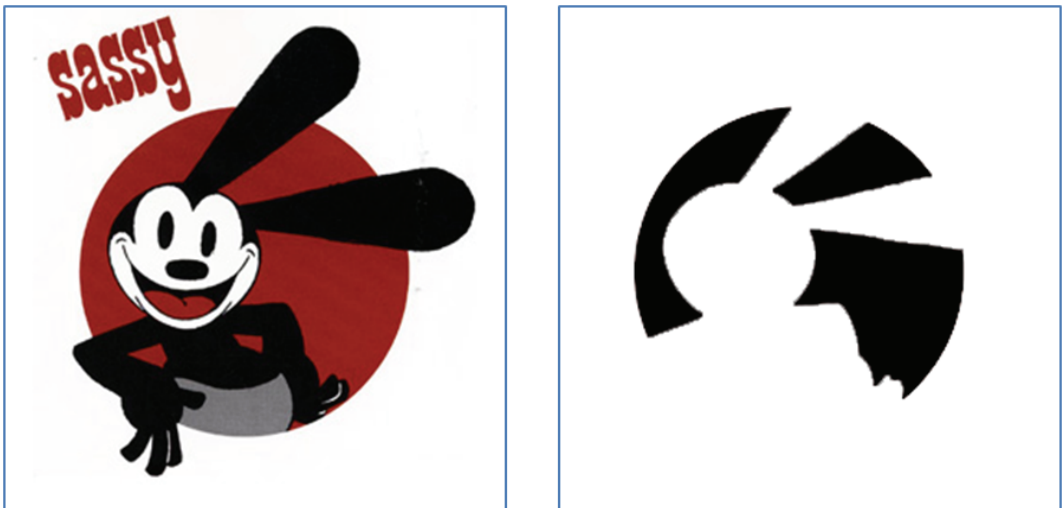
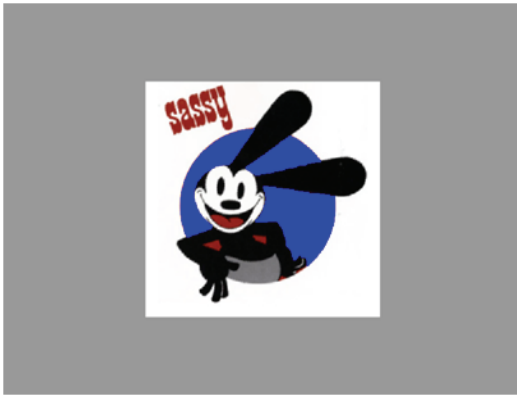


그림 37. 마우스 캐릭터의 소스 이미지(좌)와 마스크 이미지(우)

그림 38은 입력 사운드에 의해 생성된 스펙트럼 결과이다. 제안하는 시스템에서 최종적으로 얻게 되는 시뮬레이션 결과 영상이 아닌 입력 사운드의 주

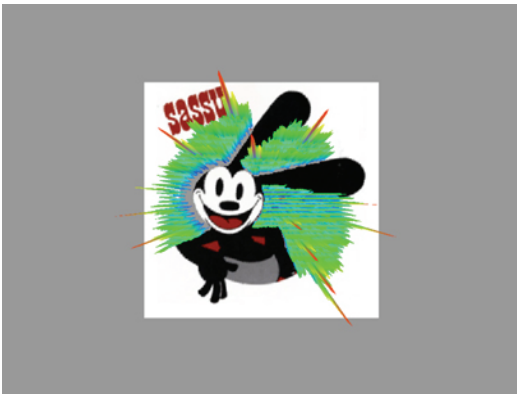
과수 성분에 따라 나타나는 스펙트럼의 성분을 분석하기 위해서이다. 그림 38의 (a)와(b)는 입력 사운드의 주파수 크기가 0일 때, 정면과 측면 결과이다. 그림 (c),(d),(e)(f)는 입력 사운드의 주파수 값에 따라 달라지는 결과를 보여준다. (c)와 (d)에서의 주파수 값 보다 (e)와 (d)에서의 값이 더욱 크기 때문에 스펙트럼 또한 높은 대역을 포함하는 것을 알 수 있다.



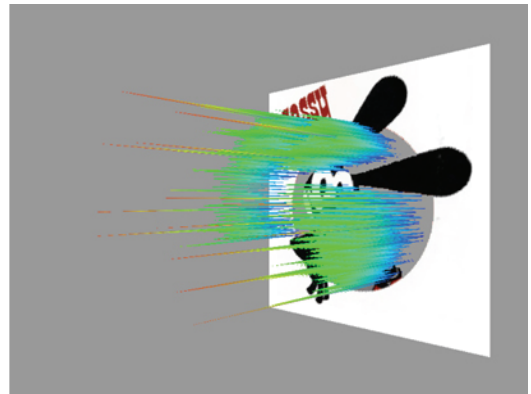
(a)



(b)



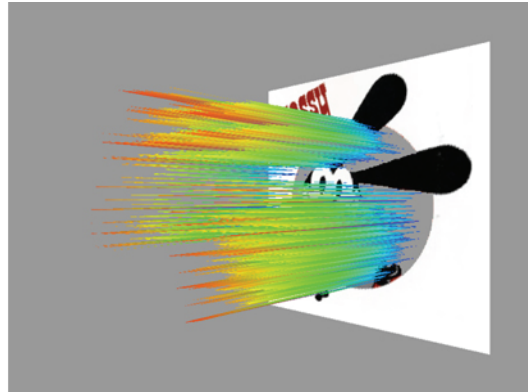
(c)



(d)



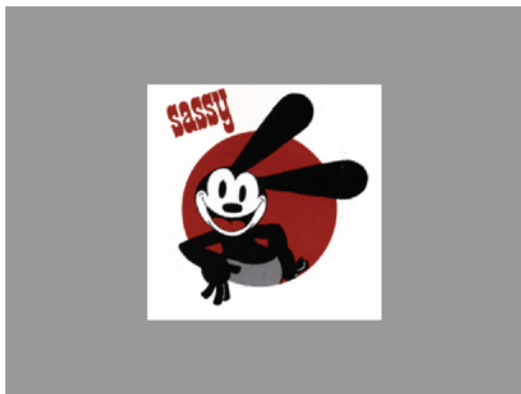
(e)



(f)

그림 38. 주파수 값 분석을 위한 스펙트럼 결과

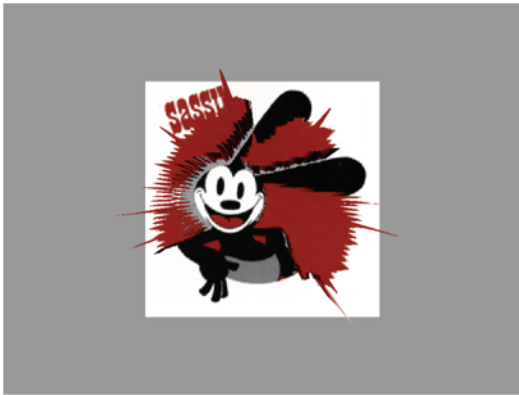
그림 39는 그림 37을 입력 이미지로 해서 본 시스템에 의해 생성되는 최종 결과 영상이다. 그림 38의 스펙트럼 값에 따라 이퀄라이저의 크기도 달라지는 것을 확인할 수 있으며, 관심영역으로 지정된 배경부분에서 비주얼 이펙트가 나타나고 있는 것을 볼 수 있다.



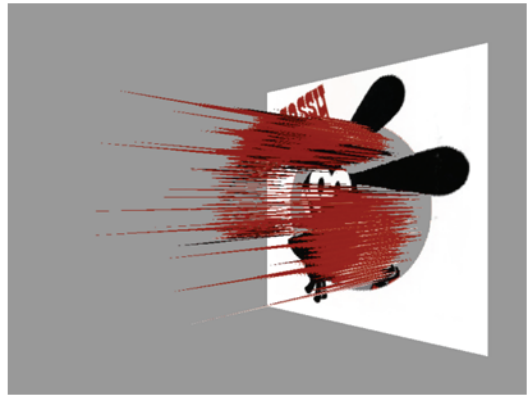
(a)



(b)



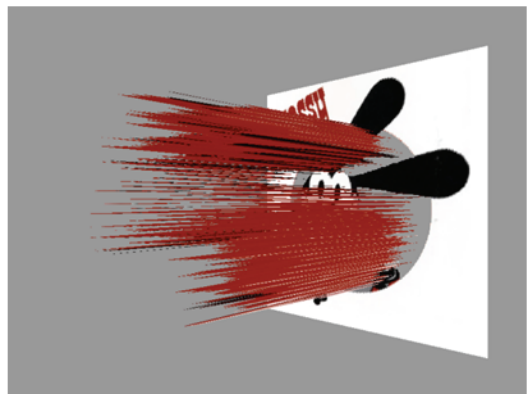
(c)



(d)



(e)



(f)

그림 39. 마우스 캐릭터의 시뮬레이션 결과

그림 40은 감성적인 시뮬레이션 결과를 얻기 위한 소스 이미지와 마스크 이미지이다. 소스 이미지에서 그 이미지를 대표하는 의미를 담고 있는 심볼인 음표를 관심영역으로 지정하였고, 입력 사운드는 잔잔한 피아노곡(Kiss the rain-이루마)을 선택하였다. 주파수 데이터는 7,663개의 픽셀에 맵핑되었다. 그림 41의 (a),(b)는 입력 사운드의 주파수 값이 0일 때의 스펙트

림을 보여주고, (c)와 (d)는 사운드가 재생될 때의 스펙트럼 값을 보여준다. 여기서 사용한 사운드는 잔잔한 피아노곡이기 때문에 스펙트럼의 최대 값이 크지 않음을 확인 할 수 있다.

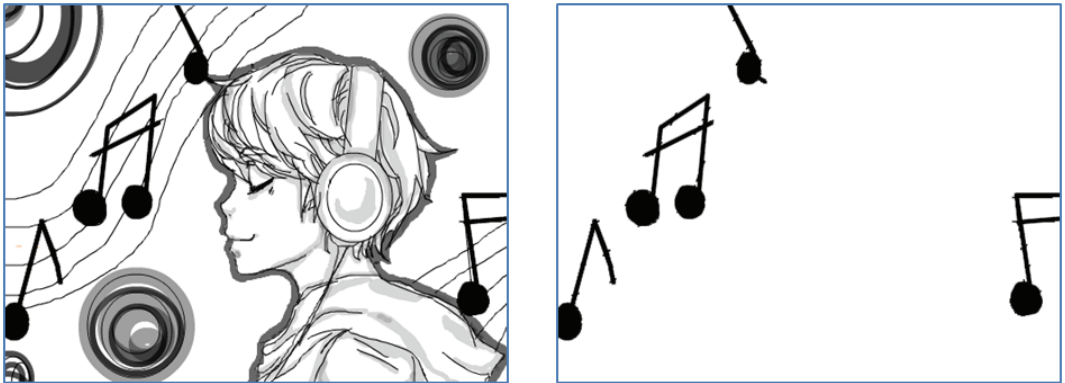
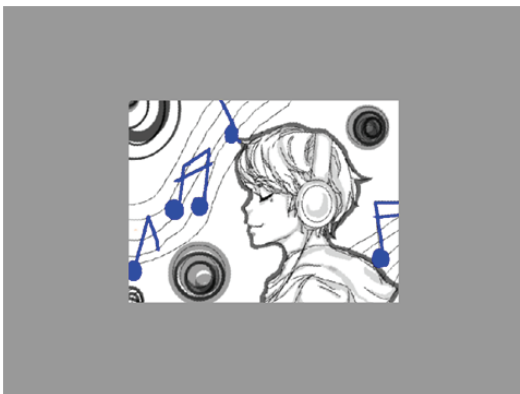
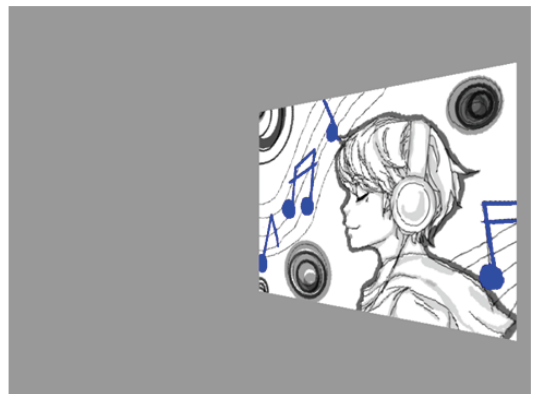


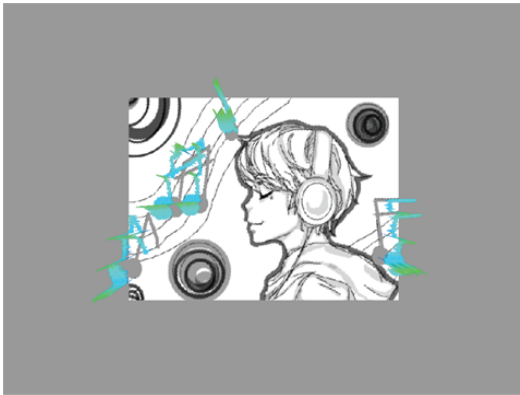
그림 40. 음표 심볼의 소스 이미지와 마스크 이미지



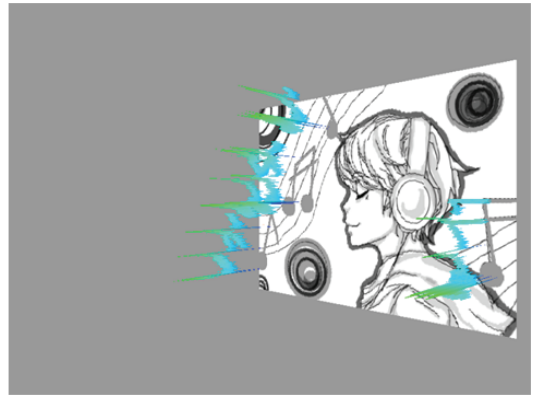
(a)



(b)



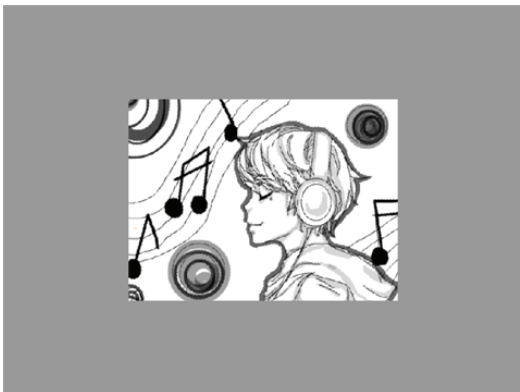
(c)



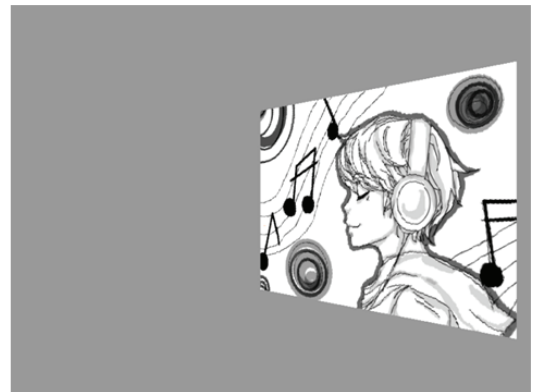
(d)

그림 41. 그림 40의 주파수 값 분석을 위한 스펙트럼 결과

그림 42는 음표 시뮬레이션의 최종 결과 영상이다. 그림 41에서 보여주는 스펙트럼의 크기와 같은 높이를 가지는 이퀄라이저가 생성되었고, 음표의 형태를 잘 유지하고 있는 것을 볼 수 있다.



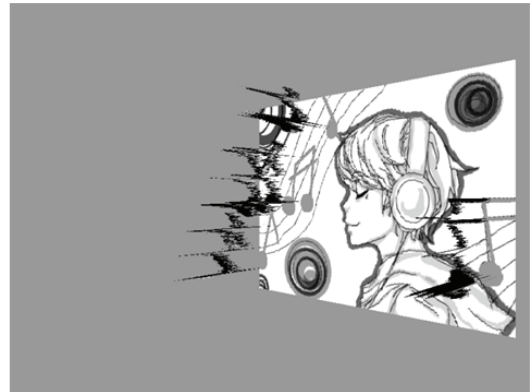
(a)



(b)



(c)



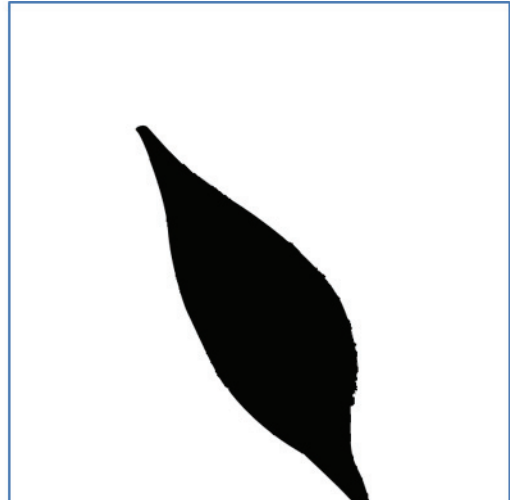
(d)

그림 42. 음표 이미지의 시뮬레이션 결과

다음은 나뭇잎 이미지를 이용한 시뮬레이션 결과이다. 그림 43의 소스 이미지와 마스크 이미지를 사용하였고, 512개의 주파수 데이터는 21,040개의 픽셀에 맵핑되었다. 적은 개수의 주파수 데이터가 상대적으로 많은 픽셀 데이터에 맵핑되었음에도 나뭇잎의 형태를 유지하는 시뮬레이션 결과를 얻을 수 있었다(그림 44).



(a)



(b)

그림 43. 나뭇잎 소스 이미지(좌)와 마스크 이미지(우)



그림 44. 나뭇잎 시물레이션의 정면(좌)과 측면(우) 결과

다음은 스머프 캐릭터와 총소리 사운드를 사용한 시물레이션 결과이다. 소스 이미지는 그림 45의 왼쪽 이미지를 사용하였으며, 관심영역은 파파 스머프의 모자와 신발을 지정하였다. 주파수 데이터는 5,157개의 픽셀에 맵핑되었으며 시물레이션 결과는 그림 46과 같다.



그림 45. 스머프 소스 이미지(좌)와 마스크 이미지(우)



그림 46. 스머프 시뮬레이션 결과

그림 47은 스머프 시뮬레이션에서 사용한 사운드 데이터를 시간의 흐름에 따라 변화하는 주파수 값으로 나타낸 것이다. 효율적인 결과 분석을 위해 입력 사운드는 총소리를 사용하였다.

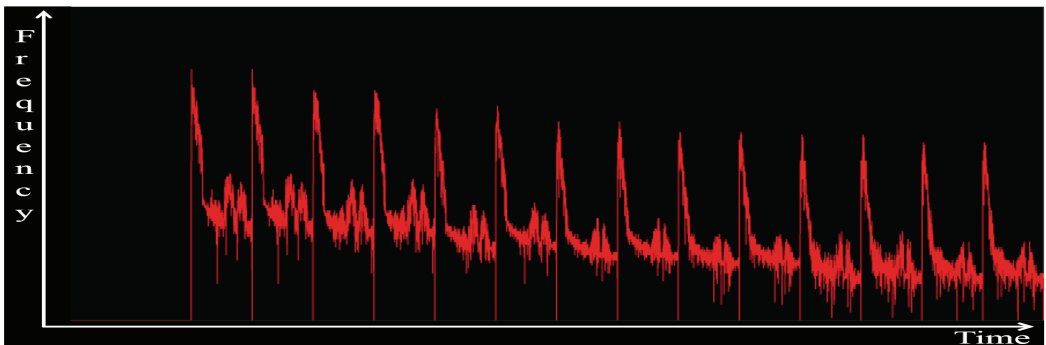


그림 47. 입력 사운드의 주파수 값 변화

그림 48은 위에서 분석한 총소리 사운드를 입력 사운드로 사용하여 렌더링 하였을 때 실제 이미지에 맵핑된 주파수 값을 나타낸다. 실제 입력 사운드의 주파수보다 훨씬 많은 픽셀 수에 맵핑하였기 때문에 그림 47과 비교하였을 때 주파수 성분이 많은 것을 알 수 있다. 하지만 전체적으로 입력 사운드에서 반복되는 패턴 즉, 주파수 특징이 잘 나타나 있는 것을 알 수 있

으며 이를 통해 입력 사운드가 가지는 무드가 잘 표현되고 있음을 알 수 있다. 그림 48에는 너무 많은 데이터가 포함되어 있기 때문에 그림 47의 특정 부분을 확대하였다(그림 49).

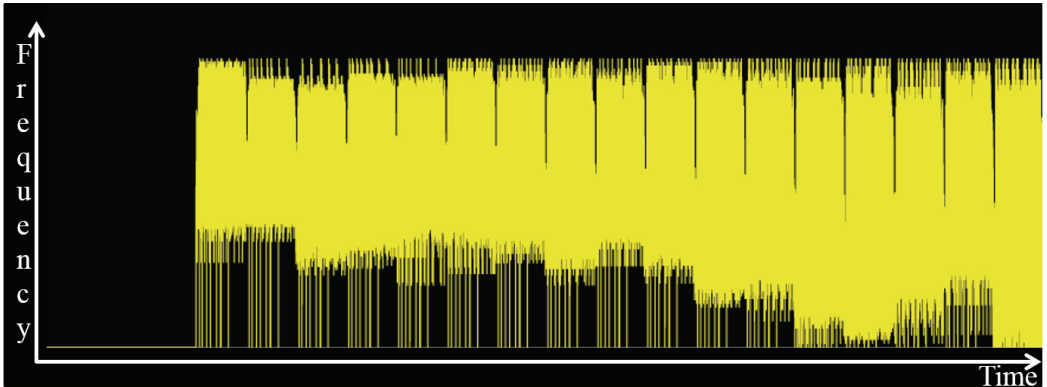


그림 48. 이미지에 맵핑된 주파수 값 변화

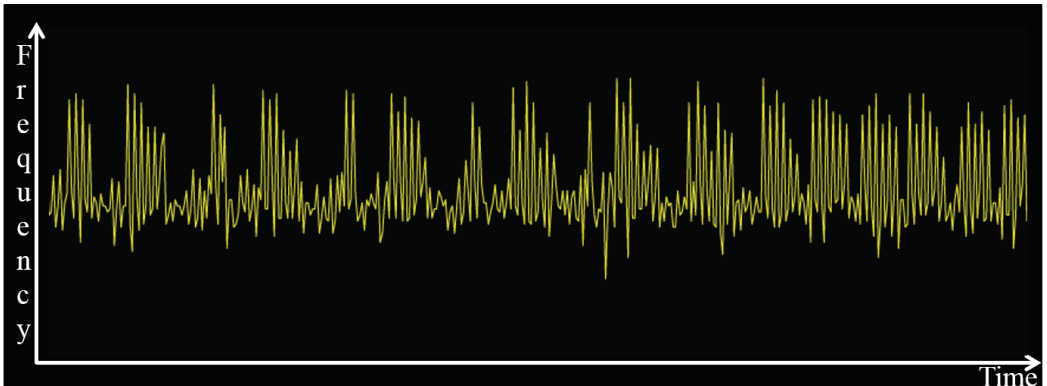


그림 49. 그림48의 부분 확대 영상

## VI. 결론 및 향후 연구

본 논문에서는 사운드 이퀄라이저 기법을 사용하여 입력 이미지로부터 이미지의 콘텐츠 형태를 유지하는 2.5D 동적 시뮬레이션을 위한 프레임워크를 제안하였다. 이미지의 콘텐츠를 유지시키기 위한 방법으로 영상 처리 분야에서 형태를 분석하기 위해 사용되는 마스킹 기법을 사용하였다. 하나의 입력 이미지로부터 여러 마스킹 이미지를 추출하였으며, 추출된 마스킹 이미지들은 서로 합성되어 또 다른 형태의 마스킹 이미지를 만들 수 있도록 하였다. 2D 이미지로부터 동적인 움직임을 만들어내는 방법으로는 이퀄라이저 기법을 사용하였다. 이퀄라이저를 생성하기 위해 우선적으로 입력 사운드를 고속 푸리에 변환하여 주파수 성분을 추출하였고, 추출된 데이터로 이퀄라이저를 생성하였다.

결과적으로는 단순히 그래프를 생성하는 기존의 이퀄라이저와는 다르게 사용자가 원하는 이미지의 콘텐츠를 유지시키면서 입력 음악과 동기화되는 이퀄라이저를 생성하였다. 최종적으로는 정적인 2차원 이미지를 동적인 2.5차원 시뮬레이션으로 표현하였으며, 이는 사용자의 피드백을 받아 이퀄라이저를 생성할 수 있다는 점에서 기존 이퀄라이저와 차별성을 가지고 있다.

현재는 사운드의 주파수 분석을 통해 사운드와 이퀄라이저를 연동하였는데 추후에 사운드의 특징으로 주파수뿐만 아니라 음악의 무드와 같이 사운드가 가지고 있는 다양한 특징들을 분석해 이퀄라이저에 적용한다면 시각적으로 더욱 다양한 효과를 생성할 수 있을 것이다. 또한 본 논문에서는 동적인 움직임을 이퀄라이저를 통해서 시뮬레이션 하였는데, 이 외에도 파티클 시뮬레이션, 유체 시뮬레이션, 버블 시뮬레이션과 같은 기술을 활용하면 다양한 사운드 시각화 결과를 만들어 낼 수 있을 것이다.

## 참 고 문 헌

- [1] Sangjick Jun, "Seeing Music by Eyes, Hearing Paintings by Ears", Journal of the Korean Society of Civil Engineers, Vol.55, No.12, pp.78-83, (2007)
- [2] Issac Newton, OPTICKS, (1704)
- [3] Wassily Kandinsky, "예술에 있어서 정신적인 것에 대하여 (Über das Geistige in der Kunst)", (1911)
- [4] 김길호, 백정기, 사운드 컬러 하모니즘: 음악이 흐르는 컬러 배색 사전, 임프레스, (2003)
- [5] KyeongA Kwon and HaeWon Byun, "Sound Feature-Driven Painterly Rendering", Korea Computer Graphics Society Conference, pp.213-214, (2012)
- [6] KyeongA Kwon and HaeWon Byun, "Sound Feature-Driven Painterly Rendering", International Journal of Multimedia and Ubiquitous Engineering, Vol.8, No.1, pp.85-90, January, (2013)
- [7] SI. Kim and JS. Jung, "A basic study on the system of converting color image into sound", KIIS Spring Conference, Vol.20, No.1, pp.251-256, (2010)
- [8] Wonseop Kim, Kyuski Chang and Taeyong Kim, "Realtime 3D Terrain Generation with Sound Data", Conference on Human-Computer Interaction(HCI '08) Korea, pp.184-189, (2008)
- [9] Seungki Kim and Woohun Lee, SOUND BOUND: making a graphic equalizer more interactive and fun, The ACM SIGCHI Conference on

Human Factors in Computing Systems(CHI '13), pp.2963-2966, (2013)

[10] Juneseok Seo, Hongsung Dae and Jinswan Park, "A study of sound graphic equalizer configuration using photo image", Conference on Human-Computer Interaction(HCI '08) Korea, pp.1280-1285, (2008)

[11] Hye Hyun Kim and Seungho Park, "A Study of Making Audio-Visual Media Art Using a Gayaguem", Korea Digital Design Council, pp.191-200, (2008)

[12] Min-Joon Yoo and In-Kwon Lee, "Creating Musical Fountain Shows," IEEE Computer Graphics and Applications, Vol.29, No.5 pp.6-13, (2009)

[13] Laewoo Kang, Tianyang Gu, Geri Gay, "Harmonic paper: interactive music interface for drawing", The ACM SIGCHI Conference on Human Factors in Computing Systems(CHI '13), pp.763-768, (2013)

[14] KyeongA Kwon and HaeWon Byun, "Real-time Visual Simulation of Sound", Korea Computer Graphics Society Conference, pp.167-168, (2013)

## ABSTRACT

# Real-time Visual Simulation of Sound

Kyeong A Kwon  
Dept. of Computer Science  
The Graduate School  
Sungshin Women' s University

In this paper, we propose a framework for the Real-time Visual Simulation which keeps the image contents, we call this 'Interest Region' , of the input image using the sound equalizer technique. The image contents are a text, a character, and something else in the image. For Visual Simulation, we have analyzed the sound and the image and matched their features accordingly to each other.

The proposed system can be roughly divided into the sound and image analysis phase and the simulation phase. In the sound and image analysis phase, we have used a Fast Fourier Transform (FFT) to extract the sound frequency from an inputted sound. We also analyzed the image to extract the interest region from the input image and had extracted the interest region by using the

masking technique.

In the simulation phase, we created the equalizer based on the extracted features. The image is represented in the pixel space of the 2-dimensional, but there is a restriction of Degree of Freedom (DOF) to represent the 3-dimensional movement. In order to solve the constraints of space, we created the equalizer by the grid based image mapping. The sampled frequency data cannot be made to a 1:1 matching and pixel values of a frequency less than the number of pixels in the content region present in the image. So, using the modular arithmetic, we created the equalizer that uses the limited number of sampled frequency data.

The equalizer that is generated is able to synchronize with the input music while maintaining the contents of the image unlike common equalizer that creates simple graphs. Finally, from the 2D static image, we represent the 2.5D dynamic simulation.