

효율적이고 사실적인 거품 사운드 생성을 위한 입자 기반 사운드 매칭과 합성

신영찬^o, 김종현^{*}

^o강남대학교 소프트웨어응용학부,

^{*}강남대학교 소프트웨어응용학부

e-mail: jonghyunkim@kangnam.ac.kr

Particle-Based Sound Matching and Synthesis for Efficient and Realistic Foam Sound Generation

YoungChan Shin^o, Jong-Hyun Kim^{*}

^oSchool of Software Application, Kangnam University,

^{*}School of Software Application, Kangnam University

● 요약 ●

본 논문에서는 거품 입자의 물리적 속성을 활용하여 가상 시뮬레이션 장면에서 맞는 거품 사운드를 합성하고 사운드의 물리적 현상을 기반으로 사운드의 크기를 효율적으로 제어할 수 있는 기법을 제안한다. 현실에서는 사운드의 근원지와 청중의 위치 관계에 따라 사운드 크기의 차이가 나타나는 것을 쉽게 관찰할 수 있다. 본 논문에서는 이 문제를 효율적으로 풀어내기 위해 복잡한 3차원 유체의 움직임을 분석하는 게 아닌, 2차원으로 투영된 입자의 유동을 분석하여 사운드를 합성하고 제어하는 방식을 소개한다. 우리의 방법은 거품 사운드의 크기를 효율적으로 조절할 수 있도록 스크린 공간에서 계산된 거품 입자의 속도와 위치를 활용하여 청중의 위치 관계 및 사운드의 방향성을 확인하고, 이를 통해 거품 사운드를 사실적으로 합성하였다.

키워드: 사운드 합성(Sound synthesis), 사운드 매칭(Sound matching), 유체 시뮬레이션(Fluid simulation), 거품 사운드(Foam sound), 사운드 제어(Sound control)

I. Introduction

가상환경의 발전 가능성이 보이기 시작하면서 중요하게 언급되는 부분은 바로 몰입형 경험이다. 사용자에게 사실적인 몰입을 제공하기 위해서는 비주얼 효과, 직관적인 상호작용을 더불어 입체적이고 사실적인 사운드를 표현해야 한다. 가상공간에서 상호작용 없이 모두 같은 사운드가 들린다면 몰입을 저하시키는 요인이 될 수 있기 때문에 가상현실콘텐츠에서 사운드는 중요한 특징 중 하나이다. 사용자에게 몰입을 전달하기 위해서는 사용자와의 상호작용을 통해 거리 혹은 위치의 상관관계를 고려하여 사실감 있는 사운드를 표현해야 한다. 이를 효율적으로 모델링하기 위해 본 논문에서는 입자 기반으로 사운드를 합성할 수 있는 방법을 제안하고자 한다. 본 논문에서는 기존에 제안되었던 거품 시뮬레이션의 사운드 합성을 활용한다[1]. 합성을 진행할 때, 군집화한 입자들의 속도, 위치 등의 물리적인 정보를 활용하여 사용자와의 위치 관계를 결정하고, 이를 통해 결과적으로 사실감 있는 사운드 합성 기법을 제안한다.

II. Preliminaries

1. Related works

사운드 생성 및 합성에서는 크게 효율성을 강조한 절차적 방식과 정확성을 강조한 물리기 반 접근법이 존재한다. 대표적인 절차적 합성 방식으로는 사운드의 공간 감각을 사용자에게 전달하기 위한 사운드 합성 기법이 있으며[2], 이 기법에서는 입자의 위치와 청중 사이의 거리를 고려해 합성되는 사운드의 진폭을 감쇠시키는 방법인 절차적 방식으로 사운드의 크기를 합성했다. 또한, 물리 기반으로 사운드가 전파되는 속도를 고려하여 청중이 들을 수 있는 시간대에 맞게 사운드 전파를 지연시켜서 합성하는 연구가 진행되었다[3].

III. The Proposed Scheme

1. Attenuation of sound

기존 연구 통해 거품 입자의 높이인 y 값을 $x-z$ 평면으로 투영시킨 후 격자 형태로 나누어 거품 입자를 군집화하고 사운드를 합성한다[1].

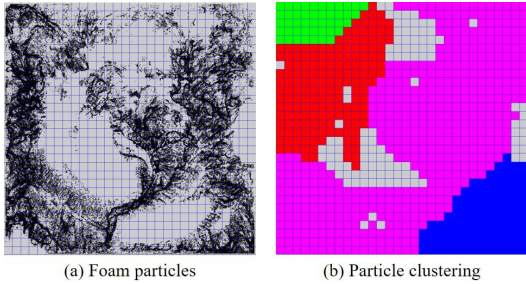


Fig. 1. Particle clustering with our method.

이때, 사운드를 합성하기 위해 기존 연구에서는 군집화한 속도 정보를 활용하여 합성을 진행했지만, 본 논문에서는 사운드의 감쇠를 적용하기 위해 군집화한 속도 정보와 위치 정보를 모두 활용한다. 우선 사운드 에너지는 모든 방향으로 퍼지기 때문에, 거리에 따라 듣는 사운드가 거의 일정하다. 하지만 사운드가 퍼지는 방향과 청중 사이의 상관관계를 고려해서 방향마다 사운드의 감쇠를 다르게 적용해야 한다. 그래서 본 논문에서는 사운드의 방향성을 계산할 때는 속도 정보를 활용하고 청중과의 위치 정보를 계산할 때는 군집화한 입자들의 무게 중심 좌표를 활용한다.

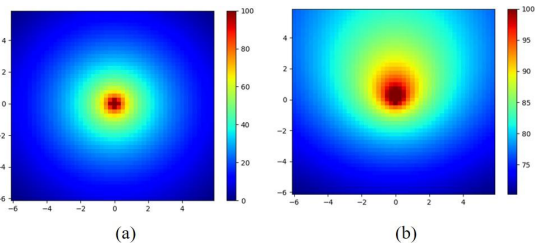


Fig. 2. Attenuation of sound with distance and direction.

Fig. 2a는 사운드의 감쇠를 모두 동일하게 적용한 것이고 Fig. 2b는 사운드의 방향성이 위쪽을 향하고 있을 때의 감쇠를 적용한 결과이다. 이때, 사운드의 감쇠를 적용하기 위한 수식은 아래와 같다 (수식 1 참조).

$$db' = db - 20 \log(r) \quad (1)$$

여기서 db' 는 사운드의 근원지에서 거리 r 만큼 떨어져 있는 사운드를 말한다. 이때, r 을 위치 관계에 따라 다르게 적용하기 위해서 아래와 같은 수식을 활용한다 (수식 2 참조).

$$r' = r \frac{\cos\theta + 2}{3} \quad (2)$$

여기서 r' 는 청중의 위치와 사운드 근원지의 중심 좌표를 활용하여 두 위치가 이루는 각도인 $\cos\theta$ 를 구하고, 그 값에 r 을 곱하여 사운드의 감쇠를 다르게 적용한다. 이때, 각도를 곱하지 않고 분모와 분자에 상수를 추가해주는 이유는 두 위치 관계가 90도를 가질 때 사운드가 소실되는 현상을 방지하기 위해서이다.

그리고 수식 1과 같이 사운드의 감쇠를 적용할 때 데시벨(db) 형태로 감쇠를 적용해야 하므로 합성되는 사운드의 진폭 값을 데시벨 형태로 변환시킨 후 감쇠를 적용하고 그 이후에 다시 진폭 값으로 변환하는 작업을 했다 (수식 3과 4 참조).

$$db = 10 \log_{10}(mag^2) \quad (3)$$

$$mag = (10^{db/10})^{0.5} \quad (4)$$

여기서 mag 는 사운드의 진폭 값을 나타낸다. 위 4개의 수식들을 시뮬레이션 장면에서 적용하면 다음과 같은 장면으로 표현된다.

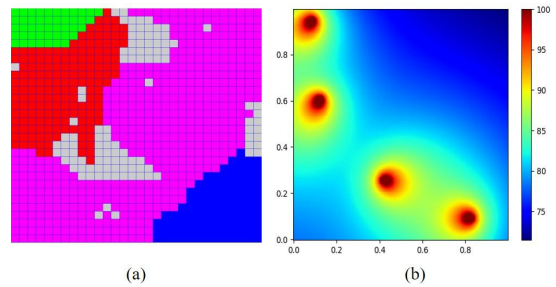


Fig. 3. Result of applying sound attenuation to the clustered scene.

Fig. 3에서 볼 수 있듯이 본 논문에서 제안하는 방법을 사용하면 군집화한 입자들의 속도와 무게 중심 좌표를 활용하여 각 군집화 영역의 사운드 방향성과 그에 따른 사운드의 감쇠를 다르게 적용할 수 있는 결과를 볼 수 있다. 수식 1-4를 시뮬레이션 적용하여 합성된 사운드의 주파수 형태는 다음과 같다 (Fig. 4 참조).

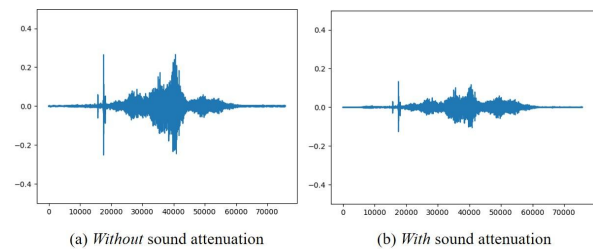


Fig. 4. Comparison of results for sound attenuation.

2. Data preprocessing

기존 연구에서는 1분가량의 파도 사운드를 0.2초씩 분할하여, 총 300개의 클립으로 나눠 합성을 진행했다[1]. 본 논문에서는 기존 연구와 합성 장면의 사운드의 정확성을 높이기 위해 20초-60초 분량의 파도 사운드 데이터 14개를 활용하였다. 이때, 14프레임마다 사운드를 탐색하도록 설정했는데, 14프레임마다 파도 사운드 데이터를 읽고 클립을 탐색해 합성하면 시간이 오래 걸린다. 그래서 매년

읽지 않게 하도록 전처리 과정에서 0.4초의 클립으로 분할하여 저장하고 저장한 파일을 활용할 수 있도록 설정해 합성 시간의 효율을 높였다 (Table 1 참조).

Table 1. Time comparison.

| | Before | After |
|------|--------|-------|
| Time | 30min | 2min |

Table 1의 경우 전처리를 하지 않았을 때는 30분이 소요되지만, 전처리한 후 합성을 진행할 때 전처리 과정까지 포함하여 2분이 소요되는 것을 확인할 수 있다.

3. Interpolation

군집화와 사운드 매핑을 통해 각 라벨에 해당하는 사운드를 추출하고, 추출된 사운드를 속도, 위치 정보를 활용하여 한 프레임에서 나오는 사운드를 하나의 사운드로 합친다. 이후 각 프레임에서 추출된 사운드를 이어 붙인다. 이때, 사운드를 단순히 이어붙이면 각 클립에서 소리의 불연속적인 부분이 발생하면서 사운드에 노이즈 구간이 생긴다. 이런 상황을 방지하기 위해 각 프레임에서 추출된 0.4초의 클립을 0.2초씩 겹쳐 보간하여 불연속적인 부분을 완화하여 사운드를 합성하였다.

4. Time delay

앞에서 언급했듯이 우리는 0.4초의 클립을 0.2초씩 겹쳐 보간하여 소리를 합성했다. 이때, 기존 연구에서는 클립의 소리가 끝나면 즉시 소리가 없어지지만, 본 논문에서는 사실감 있는 소리를 합성하기 위해 사운드가 사라지더라도 기존에 들리던 사운드는 지속될 수 있도록 조정했다. 14프레임마다 사운드 데이터를 찾고 0.2초씩 겹쳐 합성했기 때문에 소리 배열의 0.2초의 구간마다 각 클립의 평균 음량을 구하고 음량의 변화량으로 기울기를 구했다 (수식 5 참조).

$$m = \frac{\Delta y}{\Delta x} \quad (5)$$

여기서 x 의 변화량은 0.2초의 길이에 해당하는 상수로 4,456을 말하고, y 의 변화량은 현재 0.2초에 해당하는 클립의 평균 음량과 다음 클립의 평균 음량의 차이이다. 이렇게 구해진 기울기 m 을 통해 각기 다른 감쇠를 하도록 아래와 같은 커널인 F 를 사용한다 (수식 6 참조).

$$F(x) = 1 - x^m \quad (6)$$

이때, 사용되는 변수는 다음과 같다: x 는 클립이 해당하는 진폭 값, m 은 기울기이다. 이 식을 통해 기울기 값에 따라 다음과 같은 가중치 곡선을 가진다 (Fig. 5 참조).

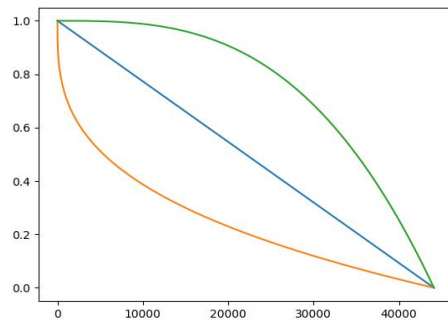


Fig. 5. The shape of the time delay function inside the kernel F .

수식 6을 통해 $m > 0$ 이라면 초록색 곡선의 형태로 사운드의 감쇠를 적용하고 $m = 1$ 이라면 파란색 곡선, $m < 0$ 이라면 주황색 곡선의 형태로 각 클립의 사운드 감쇠를 적용했다.

각 클립에서 감쇠하는 소리의 형태를 통해 소리를 찾았다면 소리 데이터 배열에서 현재 **dip**의 이후의 데이터에 소리를 더해주는 형식으로 합성을 진행한다 (Fig. 6 참조).

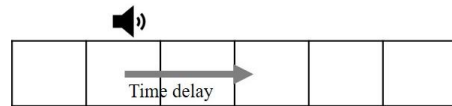


Fig. 6. Apply to one-dimensional sound array.

본 논문에서는 제안하는 Time delay 방법이 제대로 적용되는지 확인하기 위하여 소리 데이터의 1초 구간 정도의 데이터를 지운 소리와 지운 이후에 Time delay를 적용했을 때의 소리 데이터를 비교했다 (Fig. 7 참조). 이 그림을 통해서 본 논문에서 제안하는 방법이 사운드의 전파에도 영향을 주는 것을 잘 보여주고 있다. 즉, Time delay를 적용해 소리가 중간에 끊기더라도 자연스럽게 소리가 감쇠하면서 이어질 수 있도록 할 수 있다.

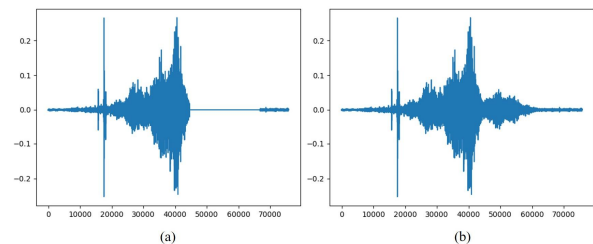


Fig. 7. Shape of sound propagation according to time delay.

Fig. 8은 본 논문에서 제안한 방법으로 합성된 거품 입자의 사운드 결과이다. 거품의 운동량에 비례해서 사운드 크기가 제어되며 거품 입자와 청중 간의 거리를 고려해 소리를 조절할 수 있었으며 사운드의 감쇠 또한 자연스럽게 이어져 합성되는 결과를 보여준다. 이 장면에서는 특정 구간에 **time delay**를 적용하지 않고, m 에 대하여 $m > 0$ 일 때에는 0.2로 클립의 주파수를 조정하여 0.4초 정도의 사운드 길이로 조정해 사운드 데이터 배열에 적용했으며, $m < 0$ 일 경우에는 0.2초

의 클립을 0.1초 정도로 조정하여 적용했다. 이렇게 조정한 이유는 기울기가 양수일 때는 이후에 사운드가 커진다는 것을 가정하여 감쇠되는 시간을 늘리고, 음수일 때는 이후 소리가 작아진다는 것을 가정해서 감쇠되는 시간을 줄였다.

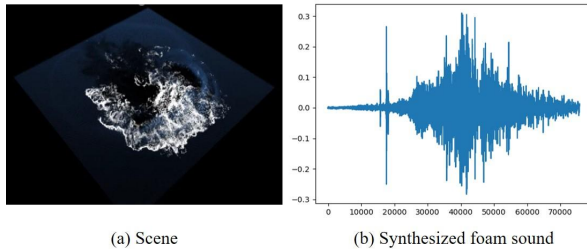


Fig. 8. Synthesized foam sound with our method.

IV. Conclusions

본 논문에서는 거품 입자의 물리적 속성을 활용하여 거품 시뮬레이션 장면에서 맞는 소리 합성 기법을 제안했다. 또한, 청중과 사운드 근원지의 거리, 그리고 사운드의 전파 방향성을 고려하고 사운드의 자연스러운 감쇠를 적용할 수 있는 새로운 방법을 제시하였다. 결과적으로 큰 계산 없이 사실적인 사운드를 효율적으로 합성해냈다.

REFERENCES

- [1] YoungChan Shin, Jong-Hyun Kim, “Efficient Foam Sound Generation with Screened Clustering Based Sound Synthesis”, Proceedings of the Summer Conference of the Korean Society for Computer Information and Information Technology Vol. 30, 2022
- [2] Su, Feng, and Chris Joslin. "Procedural sound generation for soft bodies in video games." In Motion, Interaction and Games, pp. 1-12. 2019.
- [3] James F. O'Brien, “Synthesizing sounds from physically based motion”, Computer graphics and interactive techniques, pp. 529-536. August. 2001.
- [4] Kai Wang, “Example-based synthesis for sound of ocean waves cause by bubble dynamics”, Computer animation and virtual worlds, Vol. 29, March. 2018.