

## Particle-Aware Velocity-Density Sound Mapping

Jong-Hyun Kim\*

\*Associate Professor, College of Software and Convergence (Dept. of Artificial Intelligence, Design Technology),  
Graduate School of Electrical and Computer Engineering at Inha University, Incheon, Korea

## [Abstract]

This paper proposes a data-driven sound synthesis framework that generates synchronized audio from foam particle dynamics in 3D motion videos. Energy and density descriptors derived from particle motion are mapped to audio selection, volume, and filtering. Dynamic K-means clustering and candidate reduction improve efficiency and variation, while multi-frame synthesis with cross-fading and filtering ensures temporal continuity. Experiments show responsive and coherent audio aligned with particle changes, demonstrating an interpretable approach extendable to other particle-based phenomena and learning-based models.

▶ **Key words:** Particle-Based Sound Synthesis, Velocity-Density Mapping, Fluid Simulation Audio, Dynamic Clustering, Audio-Visual Synchronization

## [요 약]

본 논문은 3D 모션 영상에서 추출한 거품 입자의 동역학을 이용하여 시간적으로 동기화된 오디오를 생성하는 데이터 기반 사운드 합성 프레임워크를 제안한다. 입자의 움직임으로부터 도출된 에너지 및 밀도 기술자를 오디오 선택, 볼륨, 필터링에 매핑한다. 동적 K-means 클러스터링과 후보군 축소를 통해 효율성과 다양성을 향상시키고, 다중 프레임 합성과 크로스페이딩 및 필터링을 통해 시간적 연속성을 보장한다. 실험 결과, 입자 변화에 대응하는 반응성과 일관성을 갖춘 오디오가 생성됨을 확인하였으며, 본 방법은 해석 가능한 접근으로서 다양한 입자 기반 현상과 학습 기반 모델로 확장 가능성을 보인다.

▶ **주제어:** 파티클 기반 사운드 합성, 속도-밀도 기반 매핑, 유체 시뮬레이션 사운드, 동적 클러스터링, 시청각 동기화

## I. Introduction

사운드 합성 기술은 가상 환경, 시뮬레이션 기반 콘텐츠, 인터랙티브 미디어에서 몰입감을 결정하는 핵심 요소 중 하나이다. 최근 컴퓨터 그래픽스 및 물리 기반 시뮬레이션 기술의 발전으로 파도, 연기, 스플래시와 같은 파티클 기반 현상을 시각적으로 매우 정밀하게 표현할 수 있게 되었으나, 이에 대응하는 오디오 생성은 여전히 수작업에 의존하거나 사전 제작된 사운드 클립을 단순히 재생하는 방식에 머무르는 경우가 많다. 이러한 접근은 시각적 변화와 청각적 반응 간의 시간적 불일치로 이어져, 복잡한 동역학을 포함하는 장면에서 현실감과 일관성을 저해하는 요인으로 작용한다.

이러한 문제를 해결하기 위해 데이터 기반 및 예제 기반 사운드 합성 기법이 제안되어 왔다. 기존 연구들은 물체의 속도, 충돌 강도, 에너지 변화와 같은 물리적 단서를 오디오 특성과 매핑함으로써, 시각적 사건에 반응하는 사운드를 생성하고자 하였다. 그러나 다수의 접근법은 제한된 동작 지표에 의존하거나, 학습 기반 모델을 활용함으로써 내부 동작의 해석 가능성이 낮고 파라미터 제어가 어렵다는 한계를 가진다. 또한 시뮬레이션 데이터를 프레임 단위로 직접 사운드에 매핑할 경우, 급격한 물리량 변화로 인해 잡음이나 청각적 불연속이 발생하는 문제가 보고되고 있다.

파티클 기반 시뮬레이션 데이터는 시각적 동작과 사운드를 연결하는 중간 표현으로서 높은 잠재력을 가진다. 특히 유체 시뮬레이션에서 생성되는 거품(foam) 또는 버블 파티클은 표면 교란, 난류 강도, 물체와의 상호작용 등 장면의 물리적 상태를 집약적으로 반영한다. 이러한 특성에도 불구하고, 거품 파티클의 속성-예를 들어 파티클 개수, 속도, 가속도-를 체계적으로 분석하고 이를 사운드 합성 규칙으로 활용한 연구는 상대적으로 부족한 실정이다.

본 논문에서는 3차원 모션 영상으로부터 추출된 거품 파티클의 물리량을 입력으로 사용하여, 장면 동역학과 시간적으로 동기화된 사운드를 생성하는 데이터 기반 사운드 합성 프레임워크를 제안한다. 제안 방법은 파티클의 속도, 가속도, 개수를 이용해 장면의 에너지 및 밀도 지표를 계산하고, 이를 오디오 클립 선택, 볼륨 제어, 필터링 단계에 순차적으로 매핑한다. 대규모 파티클 데이터의 효율적인 처리를 위해 동적 K-means 클러스터링을 적용하며, 파티클 수 기반 후보군 축소 이후 속도 및 가속도 기반의 정밀 매칭을 수행함으로써 반복적인 사운드 사용을 억제하고 장면별 변화를 효과적으로 반영한다. 또한 프레임 단위 합성에서 발생하는 잡음과 시간적 불연속을 완화하기

위해 다중 프레임 기반 합성, 크로스페이드, 로우패스 필터링을 결합한 오디오 생성 과정을 도입한다.

제안 프레임워크는 파도 및 물체-유체 상호작용을 포함한 다양한 시나리오에 대해 실험적으로 검증되었으며, 파티클 분포 변화에 따라 합성된 사운드가 동적으로 반응하면서도 청각적 연속성을 유지함을 확인하였다. 본 연구는 물리 기반 시뮬레이션 데이터와 사운드 합성 간의 해석 가능한 규칙 매핑을 제시함으로써, 학습 기반 접근법을 보완하는 실용적이고 확장 가능한 대안을 제공한다는 점에서 의미를 가진다.

## II. Related Work

### 1. Physics-Based Sound Synthesis and Rendering Pipelines

컴퓨터 그래픽스에서 사운드를 “렌더링”의 대상으로 정식화한 초기 연구로 Takala와 Hahn의 Sound Rendering은 장면의 객체/이벤트와 사운드 생성 규칙을 연결하는 파이프라인 관점을 제시하였다[1]. 이후 물리 기반 애니메이션 결과로부터 사운드를 자동 생성하려는 연구들이 등장했는데, 대표적으로 O'Brien 등은 변형/강체 운동 시뮬레이션의 동역학 정보를 이용해 충돌-접촉에 의해 발생하는 소리를 합성하는 방법을 제안하였다[2]. 또한 van den Doel 등은 인터랙티브 시뮬레이션에서 접촉력 모델과 모달(modal) 기반 진동 모델을 결합하여, 충돌-슬라이딩 등 다양한 Foley 사운드를 실시간으로 생성하는 FoleyAutomatic 프레임워크를 제시함으로써 “물리량에서 사운드” 매핑의 실용적 기반을 마련했다[3].

이와 병행하여, 물체의 진동 모드 및 방사(radiation)를 더 정밀하게 반영하기 위한 연구도 진행되었다. James 등은 복잡한 기하를 갖는 물체의 음향 방사를 실시간에 가깝게 재현하기 위해 사전계산 기반의 acoustic transfer를 도입한 Precomputed Acoustic Transfer를 제안하였다[4]. 이후 Li 등은 모달 사운드의 방사 특성을 사용자 상호작용 및 편집 관점에서 효율적으로 근사하는 방법을 제시하여, 현실적 음향 생성과 제어 가능성 간의 균형을 확장하였다[5]. 한편 접촉면의 텍스처(미세 요철)까지 고려한 접촉 사운드 합성으로 Ren 등은 textured model 간의 슬라이딩/접촉음을 물리 기반으로 합성하는 기법을 제안하여, “접촉 조건의 세밀한 기술”이 청각적 설득력에 중요함을 보여주었다[6]. 대규모 환경에서 다수의 음원과 전파를 동시에 다루기 위한 시스템 관점에서도 Raghuvanshi 등

은 대규모 장면의 인터랙티브 사운드 합성을 위한 접근을 제안하며, 실시간성·스케일러빌리티의 중요성을 강조하였다[7].

## 2. Fluid and Foam-Based Sound Synthesis

유체 장면의 사운드는 강제 충돌음과 달리 연속적인 난류, 공기 혼입, 거품의 생성/붕괴 등 복합적인 현상이 관여한다. Moss 등은 유체 시뮬레이션 결과로부터 액체 사운드를 직접 합성하는 접근을 제시하여, “시각적 유체 시뮬레이션 데이터로부터 청각적 신호” 연결 가능성을 보여 주었다[8]. 이후 거품/버블의 음향 물리를 보다 직접적으로 반영하려는 연구도 등장했는데, Toward animating water with complex acoustic bubbles는 2상(two-phase) 유체 시뮬레이션으로부터 버블 사운드를 물리 기반으로 생성하는 방향을 탐구하였다[9]. 또한 Stanford 계열 연구로 알려진 Improved Water Sound Synthesis using Coupled Bubbles는 버블 간 결합(coupling) 효과가 물 사운드의 풍부한 스펙트럼 형성에 핵심적임을 보이고, 이를 반영한 합성 프레임워크를 제안하였다[10].

본 연구 주제와 직접적으로 맞는 예제 기반 관점에서는 Wang과 Liu가 버블 다이내믹스를 기반으로 파도 사운드를 예제 기반으로 합성하는 방법을 제시하였다[11]. 이 계열 연구는 “거품/버블”이 파도 음향을 설명하는 핵심 요인임을 실증적으로 보여준다는 점에서 중요하다. 다만 일반적으로는 버블 추출·속성 산출·예제 매칭이라는 구조를 취하기 때문에, 파티클 규모가 커질수록 효율적 처리(예: 클러스터링/샘플링)와 시간적 연속성 확보가 주요 이슈로 남는다.

## 3. Data-Driven and Example-Based Mapping with Repetition Suppression

물리 기반 모델이 고정밀을 제공하는 반면, 콘텐츠 제작 관점에서는 예제 기반 및 데이터 기반 매핑이 구현 용이성과 품질·성능 절충에서 강점을 갖는다. 예제 기반 접근은 “입력 특징량(feature)로부터 오디오 조각(example)” 매칭과 이어붙이기(연결, 크로스페이드 등)로 구성되며, 반복적 클립 사용을 억제하는 전략이 청각적 자연스러움에 직접적으로 영향을 준다. 파도 사운드 예제 기반 합성 연구 역시 이러한 구조를 기반으로 하며, 버블 특성 변화에 대응하는 사운드 생성 가능성을 제시한다[11].

## 4. Learning-Based Audio-Visual Generation and Hybrid Approaches

최근에는 영상만으로 오디오를 생성하는 학습 기반 접근도 활발하다. Owens 등은 Visually Indicated Sounds에서 무음 영상으로부터 타격/스크래치 소리를 예측하고 예제 기반 합성으로 파형을 구성하는 방법을 제안하여, 시각 신호로부터 “그럴듯한” 소리를 생성할 수 있음을 보였다[12]. 또한 모달 사운드 합성 분야에서도 NeuralSound와 같은 학습 기반 방법이 제안되어, 진동 해석 및 방사 모델을 신경망과 결합해 효율성과 품질을 동시에 추구한다[13]. 다만 이러한 학습 기반 접근은 (i) 학습 데이터 편향, (ii) 물리량과 사운드 간 인과 해석의 불투명성, (iii) 장면/시뮬레이션 파라미터 변경 시 일반화 문제 등이 존재할 수 있다. 따라서 물리 기반 또는 규칙 기반 매핑을 통해 해석 가능성과 제어성을 확보하면서, 필요 시 학습 기반 요소를 결합하는 하이브리드 전략이 실용적 대안이 될 수 있다.

## 5. Positioning of the Proposed Approach

정리하면, 기존 연구들은 (a) 강제/접촉 중심의 물리 기반 사운드 합성, (b) 유체/버블 물리 기반 음향 모델링, (c) 예제 기반 매핑 및 연결 기반 합성, (d) 학습 기반 오디오 생성으로 축을 이룬다. 본 연구는 이 중 거품(foam) 파티클의 물리량(속도·가속도·개수)이라는 “해석 가능한 중간 표현”을 사용하되, 대규모 파티클을 동적으로 클러스터링하고 다단계 매칭으로 예제(클립) 선택을 안정화하며, 다중 프레임 합성 및 크로스페이드/로우패스 처리로 시간적 불연속을 완화하는 설계를 통해, 예제 기반 접근의 실용성을 강화하는 방향에 초점을 둔다.

## III. Proposed Interface Design

예제로는 총 다섯 개의 영상을 대상으로, 서로 다른 세 종류의 원본 오디오를 기반으로 사운드 매핑을 수행하였다. 또한 데이터 분석과 시각화를 위해 속도 합산 값, 파티클 수, 속도 기반 매칭 결과, 필터 적용 과정, 그리고 최종 오디오의 볼륨 변화 그래프를 각각 구현하였다.

본 연구에서는 서로 다른 유체 상호작용 특성을 갖는 5개의 시뮬레이션 영상(파도 생성, 반사, 물체 상호작용, 회전 운동 등)을 사용하여 실험을 수행하였다. 또한 제안된 사운드 매핑 기법의 일반화 성능을 평가하기 위해 3개의 서로 다른 원본 오디오 클립을 준비하고, 일부 시뮬레이션 영상에서는 동일한 오디오를 재사용하였다.

특히 동일한 오디오를 여러 영상에 적용한 실험은, 입력 오디오의 차이가 아닌 파티클 동역학의 변화에 의해 합성 결과가 어떻게 달라지는지를 분석하기 위한 목적이다. 이를 통해 제안 방법이 특정 오디오 데이터에 종속되지 않고, 다양한 물리적 장면 변화에 대해 일관된 반응을 보이는지를 검증하였다.

또한 5개의 시뮬레이션 시나리오는 파티클 밀도의 변화, 충돌에 의한 난류, 지속적인 생성 등 서로 다른 동적 특성을 포함하도록 구성되었으며, 이는 다양한 조건에서 제안 기법의 안정성과 강건성을 평가하기 위함이다.

### 1. Data Fragmentation

원본 데이터를 가공하기 위해서는 먼저 분할 과정이 필요하다. 이를 위해 원본 오디오와 원본 비디오가 각각 준비되어야 한다. 원본 오디오는 8,912개 샘플 단위로 분할해 데이터베이스를 구축한다. 한편 원본 비디오에서는 파티클 데이터의 특성을 추출하고, 오디오 길이를 조절하기 위한 시간 정보를 함께 획득한다. 데이터 매핑에 앞서 원본 오디오의 품질 또한 중요한 요소이다. 본 연구에서는 원본 오디오를 선정하기 위한 기준을 총 네 가지로 정의하였으며, 이에 대한 자세한 내용은 다음과 같다 : 1) 비디오와 어울리는 사운드, 2) 잡음이 적은 고품질 데이터, 3) 높낮이의 분포가 있는 리듬감, 4) 1분 정도의 적당한 길이.

본 논문에서 사용한 원본 오디오의 사운드 크기는 다음과 같다 (Fig. 1 참조).

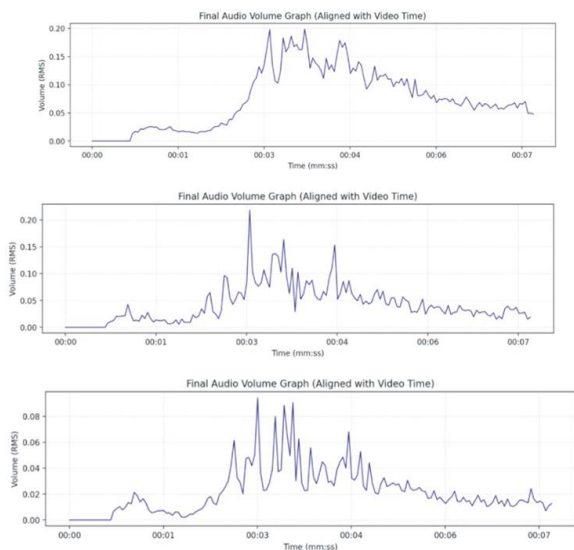


Fig. 1. RMS volume variations of different input audio signals used in the synthesis process, aligned with the video time axis.

### 2. Particle Data and Clustering

파티클 데이터를 분석할 때는 X축과 Z축만 사용하고, Y축은 제외한다. 대부분의 파티클 움직임이 수평면에서 발생하며, 수직 방향의 변화는 중력의 영향이 크게 작용하기 때문이다. 이처럼 2차원 평면 기준으로 처리하면 연산량을 약 1/3 수준으로 줄일 수 있다는 장점이 있다.

시뮬레이션 파일에는 대규모 파티클이 포함되므로, 계산 비용을 절감하고 대표 운동 특성을 안정적으로 추출하기 위해 클러스터링을 적용한다. 본 연구에서는 K-means 알고리즘으로 파티클을 K개의 군집으로 분할한 뒤, 각 군집의 평균 속도를 집계하여 시간에 따른 운동 강도를 추정한다. 특히 파티클 수  $N(t)$ 의 변화에 따라 K를 동적으로 갱신하면, 군집 분해의 해상도가 장면 밀도에 적응하면서 과도한 분산 또는 과소 대표 문제를 완화할 수 있다. 그 결과, 군집 기반 속도 집계 곡선이 전체 속도 합(velocity sum)의 시간적 추세를 더 잘 추종하며, 오디오 매핑 단계에서 장면 변화에 대한 반응성과 시간적 일관성이 향상된다(Fig. 2 참조).

본 연구에서는 클러스터 수 K를 고정값으로 설정하지 않고, 각 프레임의 파티클 수  $N(t)$ 에 따라 동적으로 결정하였다. 구체적으로, K는 파티클 수의 제곱근에 비례하는 형태로 설정하여, 계산 효율성과 대표성 간의 균형을 유지하였다. 이를 통해 파티클이 적은 구간에서는 안정적인 군집 구성이 가능하며, 밀도가 높은 구간에서는 보다 세밀한 운동 특성을 반영할 수 있다.

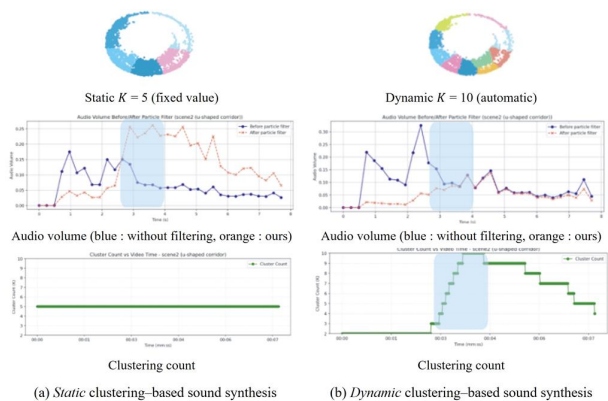


Fig. 2. Static vs. dynamic clustering-based sound synthesis. Static clustering uses a fixed  $K=5$ , whereas dynamic clustering automatically adjusts  $K$  based on particle distributions, leading to adaptive audio volume modulation.

### 3. Matching and Synthesizing

파티클 수를 기준으로 후보 오디오 그룹을 1차로 선택한다. 이후 각 클러스터링 그룹의 평균 속도를 합산한

sum\_velocity 값을 사용해, 그룹 내에서 세부 후보를 2차로 좁힌다. 마지막으로 가속도 정보를 반영하여 최종 오디오 클립을 결정한다. 이 과정에서는 최솟값과 최댓값을 기준으로 값을 정규화한 뒤, 분포 특성에 따라 오디오 매핑을 수행한다. 파티클 수나 속도가 0인 구간에는 무음을 삽입한다. 파티클 수가 많고 속도가 클수록 볼륨은 커지며, 반대로 파티클 수가 적고 속도가 작을수록 볼륨은 작아진다. 또한 동일한 소리가 반복되는 것을 방지하기 위해, 최근 사용된 30개의 클립은 재사용하지 않는다.

사운드 클립을 매핑한 뒤에는 크로스페이드(cross-fade)를 적용해 클립 간 전환이 자연스럽게 연결한다. 또한 음질을 개선하기 위해 세 가지 필터링 기법을 제안한다. 첫째, 로우패스 필터(low-pass filter)를 적용하면 급격한 변화가 완화되어 잡음이 감소하고 보다 고품질의 사운드를 생성할 수 있다. 둘째, 볼륨 스케일링(volume scaling)을 사용한다. 파티클 수와 속도를 고려해 볼륨을 비선형적으로 조절하면 장면에 따라 더욱 극적인 리듬감을 부여할 수 있다. 셋째, 파티클 수 기반 가중치를 적용한다. 파티클 수는 전체 움직임의 에너지와 밀도를 좌우하므로, 집단적 움직임의 특성을 사운드에 반영하는데 효과적이다. 이를 통해 시청각적 흐름이 보다 일관되고 안정적인 결과를 얻을 수 있다. 한편 매 프레임마다 합성할 경우 클립 1개의 길이가 지나치게 짧아져 노이즈가 발생할 수 있으므로, 14프레임 단위로 합성하는 방식이 가장 적절하다. 이는 분할된 단일 클립 길이(8,912 샘플)에 맞춰 파라미터를 조정하며 도출한 결과이다. 최종 오디오 길이는 원본 영상 시간에 맞추어 정렬하며, 크로스페이드 길이 역시 동적으로 설정해 시간 오차를 최소화한다.

## IV. Experimental Results

본 절에서는 제안한 사운드 합성 프레임워크의 실제 구현 과정을 설명하고, 서로 다른 파티클 생성 양상과 운동 특성을 갖는 다섯 가지 대표적인 시뮬레이션 예제를 통해 시스템의 동작을 분석한다. 각 예제는 파도의 형성, 반사, 물체 상호작용, 회전 운동 등 다양한 장면 조건을 포함하며, 속도 기반 매칭과 파티클 수 가중치 기법이 이러한 조건에서 어떻게 작동하는지를 단계적으로 확인할 수 있도록 구성되었다. 이후 각 하위 절에서는 Beginning of the Wave, Reflected Wave, Boxes-Created Wave, Sphere-Created Wave, Propeller-Created Wave 순으로 구현 세부 사항과 합성 결과를 제시한다.

### 1. Beginning of the Wave

첫 번째 예제에서는 속도 합이 초기부터 빠르게 증가하지만, 거품 파티클 수는 초반에는 낮게 유지되다가 후반부에 급격히 증가한다.

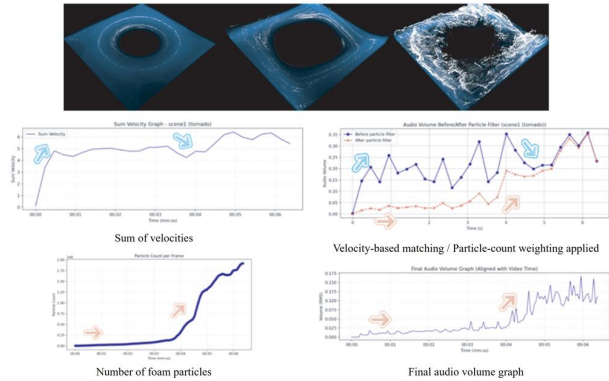


Fig. 3. Velocity- and particle-based sound synthesis results for the first example. Particle-count weighting reduces early volume overestimation and yields a final audio volume that better follows particle density evolution.

따라서 속도 기반 매칭만 적용할 경우 초기 구간의 오디오 볼륨이 과대 추정되는 경향이 나타나며, 파티클 수 기반 가중치(필터)를 함께 적용하면 입자 밀도가 충분히 형성되기 전 구간의 볼륨이 억제된다. 결과적으로 파티클 수 증가가 본격화되는 시점 이후에는 볼륨이 점진적으로 확장되며, 최종 RMS 볼륨 곡선이 장면의 물리적 변화(속도·밀도)를 더 일관되게 반영한다 (Fig. 3 참조).

### 2. Reflected Wave

두 번째 예제에서는 U자형 복도에서 유체가 벽과 두 차례 충돌하면서 국소적으로 큰 속도 피크가 발생한다. 속도 합 그래프는 충돌 시점에서 두 번의 뚜렷한 최대값을 보이지만, 거품 파티클 수는 초기에는 거의 0에 가깝게 유지되다가 충돌 이후 난류와 혼입이 누적되면서 일정 시간 지연을 두고 급격히 증가한다. 따라서 속도 기반 매칭만 적용하면 초기 충돌 구간에서 오디오 볼륨이 과대하게 형성될 수 있으나, 파티클 수 가중치를 함께 적용하면 거품이 충분히 생성되기 전의 구간이 억제되어 초반 볼륨이 상대적으로 낮게 유지된다. 이후 파티클 수가 증가하는 구간에서는 볼륨이 점진적으로 상승하며, 최종 RMS 볼륨 곡선은 '충돌 이벤트(속도 피크)'와 '거품 생성량(파티클 수)'을 동시에 반영한 형태로 정렬된다 (Fig. 4 참조).

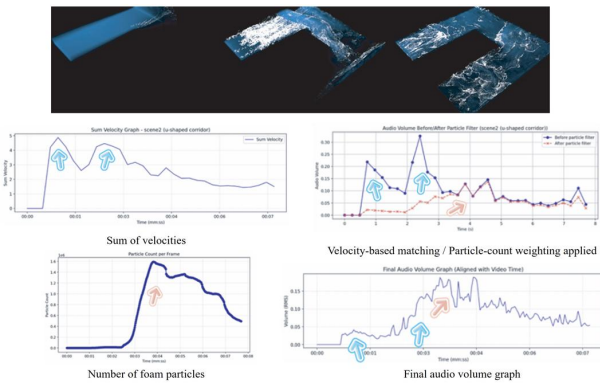


Fig. 4. Velocity- and particle-based sound synthesis for the wall-collision scenario.

### 3. Boxes-Created Wave

세 번째 예제에서는 두 개의 상자가 이동하면서 거품 파티클이 지속적으로 생성되며, 이에 따라 파티클 수가 초반에 급격히 증가한 뒤 일정 구간에서 높은 수준을 유지하고 후반부에 완만히 감소하는 경향을 보인다.

반면 속도 합은 초기 상승 이후 큰 변동 없이 비교적 안정적으로 유지되므로, 속도 기반 매칭만 적용할 경우 중반 이후 볼륨 변화가 제한적이며 일부 구간에서 순간적인 피크가 과대하게 반영될 수 있다. 파티클 수 가중치를 함께 적용하면, 입자 밀도 변화가 볼륨 조절에 반영되어 중반부에서는 안정적인 볼륨을 유지하면서도 파티클 수가 증가하는 구간에서는 점진적으로 상승하고, 감소하는 구간에서는 완만히 낮아지는 형태로 최종 RMS 볼륨 곡선이 정리된다. 결과적으로 최종 오디오 볼륨은 상자 이동에 의해 형성되는 파티클 생성량의 장기 추세를 따르면서, 단기적인 속도 스파이크로 인한 불필요한 변동을 완화한다 (Fig. 5 참조).

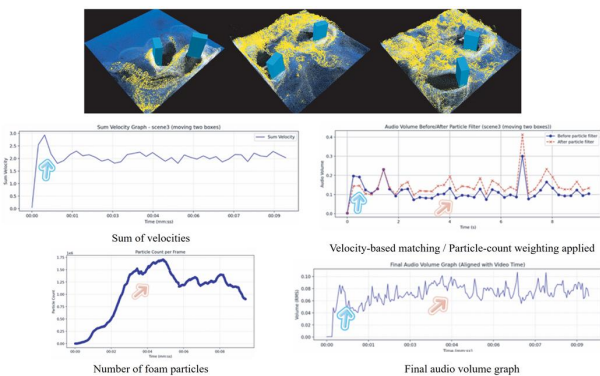


Fig. 5. Velocity- and particle-based sound synthesis for the moving two-boxes scenario.

### 4. Sphere-Created Wave

네 번째 예제에서는 회전하는 구체(스피닝 이미터)가 궤적을 따라 이동하면서 거품 파티클이 누적 생성되어, 파티클 수가 시간에 따라 단계적으로 증가하는 양상을 보인다. 반면 속도 합은 초기 구간에서 빠르게 상승한 뒤 비교적 안정적인 수준으로 유지되므로, 속도 기반 매칭만 적용할 경우 전체 구간의 볼륨 변화가 제한되거나 순간적인 피크에 민감해질 수 있다.

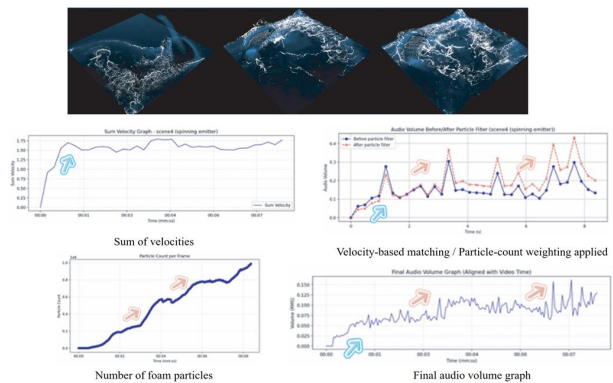


Fig. 6. Velocity- and particle-based sound synthesis for the spinning emitter scenario.

파티클 수 가중치를 함께 적용하면 누적 생성되는 거품 밀도의 변화가 볼륨 조절에 반영되어, 최종 RMS 볼륨 곡선이 점진적으로 상승하면서 후반부에는 파티클 증가 및 국소 난류 이벤트에 의해 변동성이 증가하는 패턴을 나타낸다. 결과적으로 본 예제는 ‘속도는 포화되지만 밀도는 누적되는’ 장면에서 파티클 수 기반 가중치가 장면 진행에 따른 음량 변화를 안정적으로 유도함을 보여준다 (Fig. 6 참조).

### 5. Propeller-Created Wave

다섯 번째 예제에서는 프로펠러의 회전으로 인해 거품 파티클이 지속적으로 생성·축적되면서 파티클 수가 급격히 증가한 뒤 일정 수준에서 포화되는 반면, 속도 합은 초기 피크 이후 점진적으로 감소하는 경향을 보인다. 이는 장면의 청각적 강도가 단순히 순간 속도 크기만으로 결정되지 않으며, 공기 혼입과 거품의 누적(밀도)과 같은 ‘지속적 생성 효과’가 중요한 역할을 함을 시사한다. 속도 기반 매칭만 적용할 경우 초기 회전 가속 구간의 피크가 볼륨에 크게 반영되는 반면, 이후 구간에서는 속도 감소에 의해 볼륨이 과소 추정될 수 있다. 파티클 수 가중치를 함께 적용하면 거품 밀도의 증가가 볼륨 조절에 반영되어, 속도 감소 구간에서도 일정 수준의 음량이 유지되며 최종 RMS

볼륨 곡선이 장면의 '생성량 기반 에너지' 추세를 보다 일관되게 추종한다 (Fig. 7 참조).

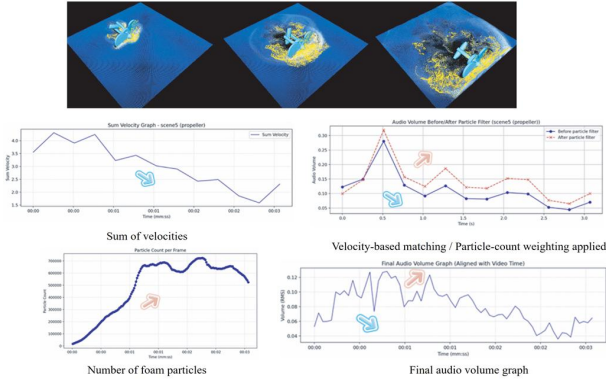


Fig. 7. Velocity- and particle-based sound synthesis for the propeller scenario.

## V. Implementation Details

본 절에서는 제안한 파티클 기반 사운드 합성 프레임워크의 구현 세부 사항을 설명한다. 시스템은 시뮬레이션 데이터로부터 파티클의 물리적 속성을 추출하고, 이를 오디오 합성에 활용하는 데이터 기반 파이프라인으로 구성된다. 전체 구현은 Python 환경에서 수행되었으며, 파티클 분석, 클러스터링, 오디오 매칭 및 합성 과정을 단계적으로 처리한다.

입력으로 사용되는 시뮬레이션 파일에는 각 프레임마다 대규모의 파티클 위치 및 속도 정보가 포함되어 있다. 본 연구에서는 계산 효율과 장면의 주요 운동 특성을 고려하여 파티클의 수평 운동 성분만을 사용하였으며, 중력의 영향이 지배적인 수직 성분은 제외하였다. 각 프레임  $t$ 에서 파티클 속도 벡터  $v_i(t)$ 를 이용해 전체 운동 강도를 나타내는 속도 합을 계산하고, 동시에 파티클 수  $N(t)$ 를 추출하여 장면의 밀도 지표로 사용한다.

파티클 수가 매우 많은 경우 개별 파티클을 모두 고려한 매칭은 비효율적이며, 본 연구에서는 K-means 클러스터링을 적용하여 파티클을 군집화한다. 정적 클러스터링에서는 클러스터 수  $K$ 를 고정값으로 설정하지만, 본 연구의 핵심 구현 요소는 파티클 수 변화에 따라  $K$ 를 동적으로 조정하는 방식이다. 이를 통해 장면 밀도가 낮은 구간에서는 과도한 분할을 방지하고, 파티클 수가 증가하는 구간에서는 보다 세밀한 대표 운동 특성을 추출할 수 있다. 각 클러스터의 평균 속도는 이후 오디오 매칭 단계에서 주요 특징량으로 사용된다.

오디오 매칭은 두 단계로 수행된다. 먼저 파티클 수  $N(t)$ 를 기준으로 후보 오디오 클립 그룹을 선택하여, 장면 밀도와 맞지 않는 사운드가 선택되는 것을 방지한다. 이후 각 클러스터의 평균 속도를 합산한 속도 합 지표를 이용해 세부 후보를 정렬하고, 가속도 변화까지 고려하여 최종 오디오 클립을 결정한다. 이 과정에서 파티클 수 기반 가중치를 적용함으로써, 속도 피크가 발생하더라도 파티클 밀도가 충분하지 않은 초기 구간에서는 볼륨이 과대 추정되지 않도록 조정한다.

선택된 오디오 클립은 프레임 단위로 직접 연결할 경우 불연속적인 잡음이 발생할 수 있으므로, 본 연구에서는 다중 프레임 단위로 오디오를 합성한다. 실험적으로 14프레임 단위의 합성이 가장 안정적인 결과를 보였으며, 이는 분할된 오디오 클립의 길이와 시간 정렬 오차를 고려하여 결정되었다. 클립 간 전환에는 크로스페이드를 적용하고, 로우패스 필터를 사용하여 고주파 잡음을 완화하였다. 최종 오디오 볼륨은 RMS 기준으로 정규화되어 비디오 시간 축과 정렬된다.

제안한 구현 방식은 서로 다른 파티클 생성 메커니즘을 갖는 다섯 가지 시뮬레이션 예제에 적용되었다. 파도의 초기 형성(Beginning of the Wave), 벽과의 반사(Reflected Wave), 이동하는 두 상자에 의한 파티클 생성(Boxes-Created Wave), 회전하는 구체에 의한 누적 생성(Sphere-Created Wave), 그리고 프로펠러 회전에 따른 지속적 거품 생성(Propeller-Created Wave)은 각각 상이한 속도-밀도 관계를 보여준다. 이를 통해 제안한 동적 클러스터링과 파티클 수 가중치 기반 매칭이 다양한 장면 조건에서도 안정적으로 동작함을 확인하였다.

다양한 프레임 길이에 대해 실험적으로 비교한 결과, 짧은 구간에서는 오디오 클립 간 전환으로 인한 노이즈가 증가하고, 긴 구간에서는 장면 변화에 대한 반응성이 저하되는 경향이 확인되었다. 그 중 14프레임 단위가 시간적 연속성과 반응성 간의 균형을 가장 잘 만족하는 설정으로 나타났다.

제안된 프레임워크의 주요 계산 비용은 K-means 기반 클러스터링 단계에서 발생하며, 일반적으로  $O(NK)$ 의 시간 복잡도를 가진다. 본 연구에서는 파티클 데이터를 수평 평면 기준으로 축소하여 연산량을 감소시키고, 파티클 수 기반 후보군 축소를 통해 오디오 매칭 단계의 탐색 공간을 제한함으로써 전체 계산 비용을 효율적으로 관리하였다.

현재 구현은 Python 환경에서 수행되었으며, 중간 규모의 파티클 데이터에 대해 안정적인 실행이 가능함을 확인하였다. 다만, 현재 시스템은 완전한 실시간 처리보다는

준실시간 수준에 해당하며, 향후 GPU 기반 병렬 처리 및 근사 클러스터링 기법을 도입하여 실시간 응용 가능성을 확장할 계획이다.

## VI. Conclusions

본 논문에서는 3차원 시뮬레이션 영상으로부터 추출된 거품(foam) 파티클의 물리적 특성을 활용하여, 장면 동역학과 시간적으로 동기화된 사운드를 생성하는 데이터 기반 사운드 합성 프레임워크를 제안하였다. 제안 방법은 파티클의 속도, 가속도, 개수와 같은 물리량을 해석 가능한 특징으로 정의하고, 이를 오디오 클립 선택과 볼륨 조절, 필터링 과정에 단계적으로 매핑함으로써 시각적 변화와 청각적 반응 간의 일관성을 확보한다.

특히 본 연구에서는 파티클 수 변화에 따라 클러스터 수를 조정하는 동적 K-means 클러스터링과, 속도 기반 매칭에 파티클 수 가중치를 결합한 매칭 전략을 제안하였다. 이를 통해 속도 피크가 존재하지만 파티클 밀도가 충분하지 않은 초기 구간에서는 볼륨 과대 추정을 억제하고, 반대로 파티클이 누적 생성되는 구간에서는 지속적인 음량 형성을 유도할 수 있음을 다양한 실험을 통해 확인하였다. 다섯 가지 서로 다른 시뮬레이션 예제에 대한 분석 결과, 제안 기법은 파도의 형성, 반사, 물체 상호작용, 회전 및 지속적 생성과 같은 다양한 파티클 동역학 조건에서도 안정적인 오디오 합성을 수행하였다.

본 연구의 접근법은 학습 기반 모델에 의존하지 않으면서도 물리적 의미가 명확한 규칙 기반 매핑을 사용함으로써, 사운드 합성 과정의 해석 가능성과 제어 용이성을 동시에 제공한다는 점에서 의의가 있다. 또한 제안한 프레임워크는 파도 장면이 국한되지 않고, 불, 연기, 폭발, 바람 등 다양한 파티클 기반 시뮬레이션 영상의 배경 사운드 생성으로 확장 적용될 수 있다.

향후 연구에서는 파티클 기반 특징과 학습 기반 오디오 모델을 결합한 하이브리드 접근을 통해 보다 정교한 사운드 표현을 도입하고, 사용자 평가나 정량적 지표를 활용한 품질 분석을 추가함으로써 시스템의 객관적 성능을 검증할 계획이다. 이를 통해 물리 기반 시뮬레이션과 사운드 합성 간의 통합적 오디오-비주얼 생성 프레임워크로 발전시키고자 한다.

본 연구에서는 예제 기반 사운드 합성의 특성상 기준 신호가 명확하지 않기 때문에 SNR과 같은 전통적인 신호 기반 정량 지표를 직접 적용하지 않았으며, 대신 시각적 변

화와 청각적 반응 간의 일관성을 중심으로 결과를 분석하였다. 또한 동일한 파티클 입력 조건에서 단순 속도 기반 매핑 방식과 비교하여 제안 기법의 효과를 확인하였으나, 물리 기반 음향 합성 기법이나 학습 기반 오디오 생성 방법과의 직접적인 비교는 각 접근법 간의 모델 구조 및 요구 조건 차이로 인해 제한적으로 수행되었다.

향후 연구에서는 MOS(Mean Opinion Score)와 같은 사용자 기반 평가를 도입하고, 대표적인 기존 방법들과의 비교 실험을 확장하여 정량적·정성적 평가를 포함한 보다 체계적인 성능 검증을 수행할 예정이다.

본 연구는 유체 시뮬레이션 기반 거품 파티클을 중심으로 실험을 수행하였으나, 제안 방법은 파티클의 속도, 가속도, 개수와 같은 일반적인 물리적 특징을 기반으로 사운드를 생성하는 구조를 가지므로, 불, 연기, 폭발 등 다양한 파티클 기반 현상으로 확장 적용이 가능하다. 다만 각 현상은 파티클 분포 및 음향 특성이 상이하므로, 이에 적합한 특징 정의 및 매핑 전략이 추가적으로 필요하다. 향후 연구에서는 다양한 파티클 기반 현상에 대한 확장 실험을 수행하여 제안 방법의 일반성과 적용 가능성을 체계적으로 검증할 예정이다.

## ACKNOWLEDGEMENT

This work was supported by the National Research Foundation of Korea(NRF) grant funded by the Korea government(MSIT) (RS-2026-25474535).

## REFERENCES

- [1] Takala, T., & Hahn, J. Sound rendering. In Proceedings of the 19th annual conference on Computer graphics and interactive techniques, pp. 211-220, 1992.
- [2] O'Brien, J. F., Cook, P. R., & ESSL, G. Synthesizing sounds from physically based motion. In Proceedings of the 28th annual conference on Computer graphics and interactive techniques, pp. 529-536, 2001.
- [3] Van Den Doel, K., Kry, P. G., & Pai, D. K. FoleyAutomatic: physically-based sound effects for interactive simulation and animation. In Proceedings of the 28th annual conference on Computer graphics and interactive techniques, pp. 537-544, 2001.
- [4] James, D. L., Barbič, J., & Pai, D. K. Precomputed acoustic transfer: output-sensitive, accurate sound generation for

- geometrically complex vibration sources. *ACM Transactions on Graphics (TOG)*, vol. 25, no. 3, pp. 987-995, 2006.
- [5] Li, D., Fei, Y., & Zheng, C. Interactive acoustic transfer approximation for modal sound. *ACM Transactions on Graphics (TOG)*, vol. 35, no. 1, pp. 1-16, 2015.
- [6] Ren, Z., Yeh, H., & Lin, M. C. Synthesizing contact sounds between textured models. In *2010 IEEE virtual reality conference (VR)*, pp. 139-146, 2010.
- [7] Raghuvanshi, N., & Lin, M. C. Interactive sound synthesis for large scale environments. In *Proceedings of the 2006 symposium on Interactive 3D graphics and games*, pp. 101-108, 2006.
- [8] Moss, W., Yeh, H., Hong, J. M., Lin, M. C., & Manocha, D. Sounding liquids: Automatic sound synthesis from fluid simulation. *ACM Transactions on Graphics (TOG)*, vol. 29, no. 3, pp. 1-13, 2010.
- [9] Langlois, T. R., Zheng, C., & James, D. L. Toward animating water with complex acoustic bubbles. *ACM Transactions on Graphics (TOG)*, vol. 35, no. 4, pp. 1-13, 2016.
- [10] Xue, K., Aronson, R. M., Wang, J. H., Langlois, T. R., & James, D. L. Improved Water Sound Synthesis using Coupled Bubbles. *ACM Transactions on Graphics (TOG)*, vol. 42, no. 4, pp. 1-13, 2023.
- [11] Wang, K., & Liu, S. Example-based synthesis for sound of ocean waves caused by bubble dynamics. *Computer animation and virtual worlds*, vol. 29, no. 3-4, pp. e1835, 2018.
- [12] Owens, A., Isola, P., McDermott, J., Torralba, A., Adelson, E. H., & Freeman, W. T. Visually indicated sounds. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2405-2413, 2016.
- [13] Jin, X., Li, S., Wang, G., & Manocha, D. Neursound: learning-based modal sound synthesis with acoustic transfer. *ACM Transactions on Graphics (TOG)*, vol. 41, no. 4, pp. 1-15, 2022.

## Authors



Jong-Hyun Kim received the B.A. degree in the Department of Digital Contents at Sejong University in 2008. He received M.S. and Ph.D. degrees in the Department of Computer Science and Engineering at Korea University,

in 2010 and 2016. Prof. Kim is an Associate Professor in the College of Software and Convergence (Dept. of Artificial Intelligence, Design Technology), Graduate School of Electrical and Computer Engineering at Inha University. His current research interests include fluid animation and virtual reality.