Haptic Force Guided Sound Synthesis in Multisensory Virtual Reality (VR) Simulation for Rigid-Fluid Interaction

Haonan Cheng* Shiguang Liu†

Division of Intelligence and Computing, Tianjin University, P.R. China

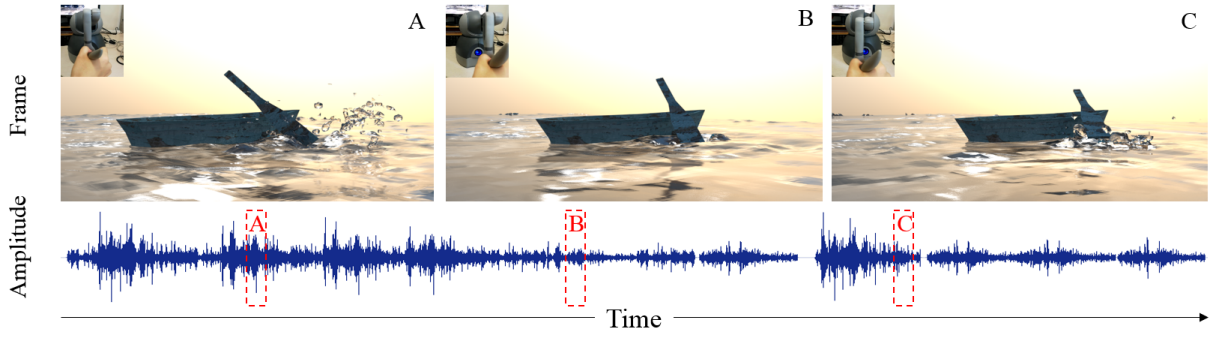


Figure 1: Illustrations of the sound results for a rigid-fluid interaction scene (boating). The top row presents three frames of the visual simulation result of the rigid-fluid interaction (from left to right are the 43th frame, the 132th frame and the 187th frame). The top left corner inset in each frame shows the corresponding haptic scene. The bottom row illustrates the waveform of the generated sound result and the red dotted boxes mark the generated audio synchronized with the visual animation.

Abstract

This paper tackles a challenging problem for interactive rigid-fluid interaction sound synthesis. One core issue of the rigid-fluid interaction in multisensory VR system is how to balance the algorithm efficiency, result authenticity and result synchronization. Since the sampling rate of audio is far greater than visual and haptic modalities, sound synthesis for a multisensory VR system is more difficult than visual simulation and haptic rendering, which still remains an open challenge until now. Therefore, this paper focuses on developing an efficient sound synthesis method tailored for a multisensory system. To improve the result authenticity while ensuring real time performance and result synchronization, we propose a novel haptic force guided granular sound synthesis method tailored for sounding in multisensory VR systems. To the best of our knowledge, this is the first step that exploits haptic force feedback from the tactile channel for guiding sound synthesis in a multisensory VR system. Specifically, we propose a modified spectral granular sound synthesis method, which can ensure real time simulation and improve the result authenticity as well. Then, to balance the algorithm efficiency and result synchronization, we design a multi-force (MF) granulation algorithm which avoids repeated analysis of fluid particle motion and thereby improves the synchronization performance. Various results show that the proposed sound synthesis method effectively overcomes the limitations of existing methods in terms of audio modality, which has great potential to provide powerful technological support for building a more immersive multisensory VR system.

Index Terms: Human-centered computing—Human computer interaction—Interaction devices—Haptic devices; Applied computing—Arts and humanities—Sound and music computing

*e-mail: tjuchn@tju.edu.cn

†Corresponding author. e-mail: lsg@tju.edu.cn

1 Introduction

A major concern of virtual reality (VR) is improving user immersion through the simulation of sensations (sight, hearing, taste, smell and touch) of human beings. Rigid-fluid interaction, as a common interaction scenario in VR, has been largely ignored. One of the main reasons is that the existing sound synthesis technology fails to provide an ideal performance. The core issue is how to balance the algorithm efficiency, result authenticity and result synchronization.

At present, there exist two classes of approaches for sound synthesis: physical based methods and non-physical based methods. The former are usually based on physical principles for sound modelling which require complex computations. Although there are several studies [3, 25, 27] which explored real time sound synthesis in multisensory VR simulation, they were concerned about rigid-rigid interaction that are not suitable for rigid-fluid interaction. This is due to that the fluid motion exhibits long-time nonlinearities which is more complicated than rigid-rigid interaction, which may increase the difficulty of sound modelling and sound synchronization. Therefore, despite several physical based methods [18, 24, 40] can synthesize quite realistic fluid sound with sound radiation models and complex acoustic bubbles, the synthesis time cannot meet real time requirement. Cirio et al. [9] adopted approximate solutions to simplify the sound model for real time rendering. This would lead to the absence of sound details, which reduces the authenticity of synthesized sound. On the other hand, the non-physical based methods utilize recordings or signal based sound textures which can set around complex sound modelling. Among them, the granular synthesis technique [11] is a typical non-physical based algorithm which can generate realistic results efficiently. However, an obvious drawback of the granular synthesis method is the unpredictable granulation, which may cause asynchronization between sound and animation. Therefore, how to balance the algorithm efficiency, sound authenticity and sound synchronization for sound synthesis of rigid-fluid interaction is one of the burning issues in multisensory VR simulation.

Fortunately, Cirio et al. [9] stated that the vibrotactile and acoustic phenomena share a common physical source. Hence, for the same phenomenon, different modalities have different expressions,

but they interact with each other. This provides the possibility of avoiding repeated computation of motion states in different channels in a multisensory VR simulation system. Based on such an idea, in this paper, we analyze the relationship between haptic and audio modality and propose a haptic force guided sound synthesis method for rigid-fluid interaction. We show an example generated by our method in Figure 1 and more results in Section 5.

Specifically, to improve the sound authenticity while ensuring real time performance, we propose a modified granular synthesis method in frequency domain which avoids the post-processing smoothing in conventional time domain methods. The complex sound is composed by simple and small acoustic grains (short sound clips) and sound phase is reconstructed to provide continuity. In order to solve the concomitant synchronization problem, we leverage the fact that vibrotactile and acoustic phenomena share a common physical source. Hence, we utilize haptic forces to guide the sound synthesis which avoid the analysis of complex motion of fluid particles. Specifically, we design a new multi-force (MF) granulation algorithm to guide the sound synthesis and the haptic forces are defined as state sequences during the granulation. In summary, the main contributions of our work are listed as follows:

- We propose a new haptic force guided sound synthesis method tailored for multisensory simulation of rigid-fluid interaction scenes. Our method can achieve more high quality sound results with competitive time performance than the state-of-the-art sound synthesis solutions for multisensory VR systems.
- To the best of our knowledge, our method is the first attempt to make use of the feedback haptic force for guiding sound synthesis in a multisensory VR system. This method can effectively avoid repeat analyzation of motion states in different channels of a multisensory VR system, and therefore greatly improves the system performance.
- A modified spectral granular synthesis method with a novel MF granulation algorithm is designed which can ensure the algorithm efficiency and sound synchronization at the same time. The quality of sound and synchronization performance has been greatly improved in comparison with previous methods.

2 Related Work

Multisensory information is essential for designing immersive virtual worlds which has received considerable attention in the VR community. In this section, we will discuss exciting techniques that are highly related to our work. Due to the interdisciplinary nature of this work, we only briefly cover a few of the most relevant existing studies in rigid-fluid simulation, haptic rendering, multisensory simulation and sound synthesis techniques.

2.1 Rigid-Fluid simulation

Realistic rigid-fluid interactions require to simulate both fluid motion and collisions between fluid and rigid bodies. For fluid motion, there are two main approaches, namely the Lagrangian viewpoint and the Eulerian viewpoint. Succinctly, the Lagrangian viewpoint corresponds to a particle system and the Eulerian viewpoint corresponds to using a fixed grid. Readers can refer to [6] for details. Although each approach has been extensively studied, we focus on the Lagrangian methods, since the simulated fluid is not bounded by a grid, which is faster to compute and suitable for multisensory VR simulation. Among Lagrangian methods, the smoothed particle hydrodynamics (SPH) method [7] is the most popular type which can effectively simulate meso-scale or small-scale free surface fluid. In order to achieve real-time fluid simulation, the GPU based SPH simulation was introduced by Amada et al. [1].

In SPH based simulations, collisions between fluid particles and rigid bodies are usually handled through representing rigid bodies as a set of particles. Some researches [4, 31] used a unified particle representation, which simulate rigid body particles as fluid particles and reduce the overall complexity of the computations. In order to achieve real time feedback, in our research, we use a SPH method for fluid simulation on GPU and the rigid body is represented as a set of unified particles. Each point in the fluid and rigid is labeled as a separate particle, with position x and velocity u.

2.2 Haptic Rendering

Since our method relies on the calculation of haptic force, here we briefly introduce haptic rendering. During the last decade, haptic has been a new emerged and interesting subject for many researches [5, 29, 30]. According to the number of degrees of freedom (DoF), the haptic rendering methods can be classified as 3-DoF haptic rendering methods and 6-DoF haptic rendering methods. Moreover, according to the interaction type, the haptic rendering methods can be classified as rigid-rigid interaction, rigid-deformable interaction and rigid-fluid interaction [37]. Yang et al. [38] used a 3-DoF device to simulate a fluid haptic scene in real time, but the 3-DoF input device narrowed the application of our research. Cirio et al. [8] firstly introduced a 6-DoF haptic rendering algorithm for rigid-fluid interaction based on SPH. Liu et al. [19] incorporated the buoyancy into haptic force calculation framework which improved the reality of the haptic force of the coupling. In our research, we choose Geomagic Touch, which is one of the most widely used haptic devices that many haptic researches [20, 22, 23] are based on. We utilize this 6-DoF input device to interact with fluid in virtual world and get the haptic force for guiding sound synthesis. For haptic rendering, we employ the SPH haptic rendering method [19] which allows to efficiently calculate several types of forces for our sound synthesis method and smooth haptic feedback.

2.3 Multisensory VR Simulation

As pointed out in [9], every time a feedback condition was added, the user immersion had been significantly improved. Thus, the *multisensory VR simulation* plays an important role in realistic perception. In this context, several researchers have demonstrated encouraging results in contact sound [3] and sound of walking on icy snow, creaking floors, brushwood [27], etc. However, the major concerns of aforementioned methods are rigid-rigid interaction. Although numerous techniques have been proposed for visual rendering [2, 15], haptic rendering [12, 39] and audio rendering [17, 28] of rigid-fluid interaction, there is a lack of researches on multisensory VR simulation. This is due to the motion analysis of fluid is more complex than the rigid body, so the calculation of haptic and sound are more complicated.

2.4 Sound synthesis

Sound synthesis methods have been developed for many years. However, sound generation algorithms of VR applications have different emphasis. Since VR applications require real time feedback, one critical issue is the synchronization between sound and other modalities that should allow an efficient mapping to ensure real time rendering. In addition, the authenticity and synchronization of sound results also need to be guaranteed, which is similar to the focus of the bimodal (visual and auditory) approaches. Here, we classify the sound synthesis methods into two categories, namely physical based method and non-physical based method. We further analyze the limitations of these two classes of methods in multisensory VR simulations.

Physical based method Recent literature [18, 24, 35, 40] have shown that sound synthesis techniques based on physical models allow for high quality synthesis and interactivity, since the physical parameters of the sound models can be naturally controlled by user gestures and actions. However, considering different types of bubble actions will result in great computation consumption. Although in [24], Moss et al. designed a simplified, physically inspired model

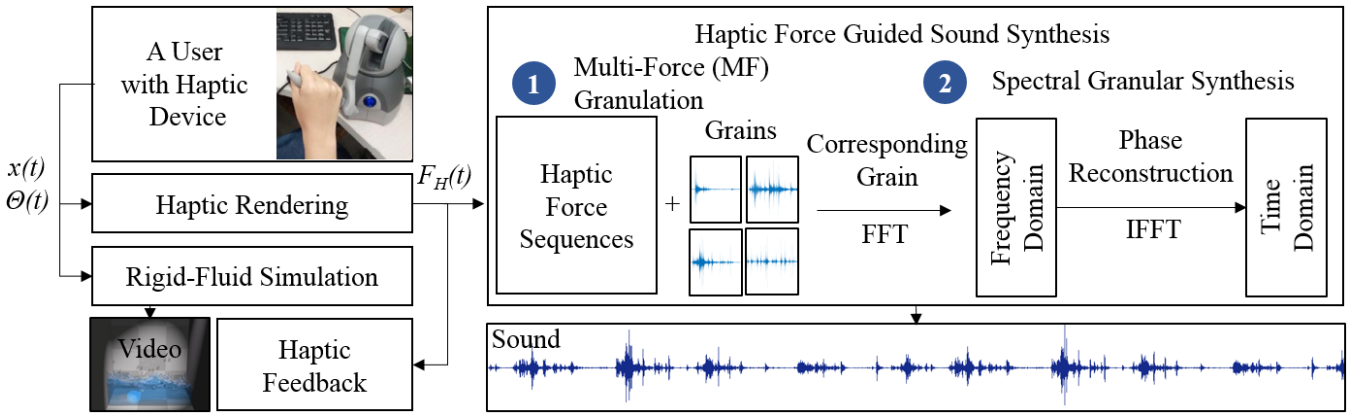


Figure 2: The architecture for our haptic force guided sound synthesis technique. In each time step, the user inputs rigid body position $x(t)$ and rotation angle $\Theta(t)$ by the haptic device to the rigid-fluid simulation module and the haptic rendering module. The force calculation module provides haptic forces $F_H(t)$ to the sound synthesis module and return the final haptic feedback. The sound synthesis module composes the grains guided by haptic force sequences based on a MF granulation algorithm and a spectral granular synthesis method. The extracted grain is transformed into frequency domain to reconstruct phase, and finally synthesizes the synchronized sound with an inverse transformation.

for real time shallow water simulation, the interaction possibilities for different scenes are reduced. Then Cirio et al. [9] proposed a physically based haptic and audio rendering model for rigid-fluid interaction. This method developed an efficient bubble generation technique based on the bubble sound model in [13] and the stochastic model in [24]. However, there are several approximate solutions in the acoustic bubble sound model, such as resonating bubble radius, initial excitation parameter, etc. These approximations can ensure real-time generation, but at the same time, they reduce the details of the sound and damage the authenticity of the sound. Thus, how to balance the efficiency of the algorithm and the authenticity of sound results becomes the technical bottleneck.

Non-physical based method Example based techniques [33, 34] are a kind of typical non-physical methods which can provide a solution of balancing computation time and sound quality. Among existing example based methods, granular synthesis [11, 32] is a relatively practical approach. In interactive environments, it is impossible to know the length needed for a sound. Because granular synthesis is a kind of overlap-add method with recorded samples, it is possible to create new sounds in any length from short examples. The granular synthesis methods can be divided into two types, namely time domain methods and frequency domain methods. Compared to time-domain grain synthesis, the methods in the frequency domain enable further manipulation of grains' sonic content, which allows to further develop timbral complexity. Thus, our method falls into the second category.

However, the granulation process of granular synthesis techniques is random which require to design an efficient mapping between parameters generated from motion and sound control parameters. Nordahl et al. [26] synthesized footstep sounds of different materials through the real footstep sounds from users detected by microphones. Wang and Liu [36] exploited a real time example-based method to synthesize the sound of ocean waves with a wave cluster method. Because the aforementioned methods are designed for their specific application scenarios, the mapping methods are not suitable for our rigid-fluid interaction scenarios. Focusing on tackling this issue, we propose a novel haptic force guided mapping algorithm to synchronize the sound and fluid motion.

3 System Overview

As sketched in Figure 2, our sound synthesis method takes haptic force F_H and grains (sound examples) as input, and outputs realistic sound which can synchronize with user interaction and visual simulations. First the haptic force F_H is calculated through a haptic rendering algorithm with the position information $x(t)$ and rotation angle $\Theta(t)$ from a haptic device (6-DoF Geomagic Touch). Then, to balance the efficiency of the algorithm and the authenticity of sound results, we propose a modified spectral granular synthesis method which can avoid the high computational cost in previous physical based method. To solve the problem of asynchronization in granular synthesis, we design a novel MF granulation algorithm to guide the process of granulation. The viscosity and buoyancy forces are sent to the sound synthesis module as guide sequences to extract the corresponding grain at each time step. Next, we transform the grains into frequency domain with fast Fourier transform (FFT), adjust the grains for enriching the diversity of results and avoiding repetitive cycles. At last, the final synchronized sound result is synthesized with inverse fast Fourier transform (IFFT). We will describe each major technical component of our system in detail in the following sections.

4 Haptic Force Guided Sound Synthesis

Our haptic force guided sound synthesis method shows great potential in multisensory VR simulation which takes advantage of introducing the information from haptic modality for improving the performance of the audio modality. As the input of our sound synthesis method, haptic force which contains pressure, viscosity and buoyancy is firstly calculated based on rigid-fluid simulation. Then, with the haptic force, we generate the grains through our designed MF granulation. Finally, we synthesize the final sound through the spectral granular synthesis based on the generated grains. To sum up, the haptic force guided sound synthesis algorithm which consists of three major parts, namely the input from physical simulation, MF granulation and spectral granular synthesis. In the following sections, we will introduce the three parts separately.

4.1 The Input of Sound Synthesis

Rigid-fluid simulation Our system is built upon incompressible Newtonian fluid whose motion is described by the Navier-Stokes equations. The SPH formulations of Navier-Stokes equations are generally expressed from momentum conservation and mass conservation as follows:

$$\frac{du}{dt} = -\frac{1}{\rho} \bigtriangledown P + \mu \bigtriangledown^2 u + F \tag{1}$$

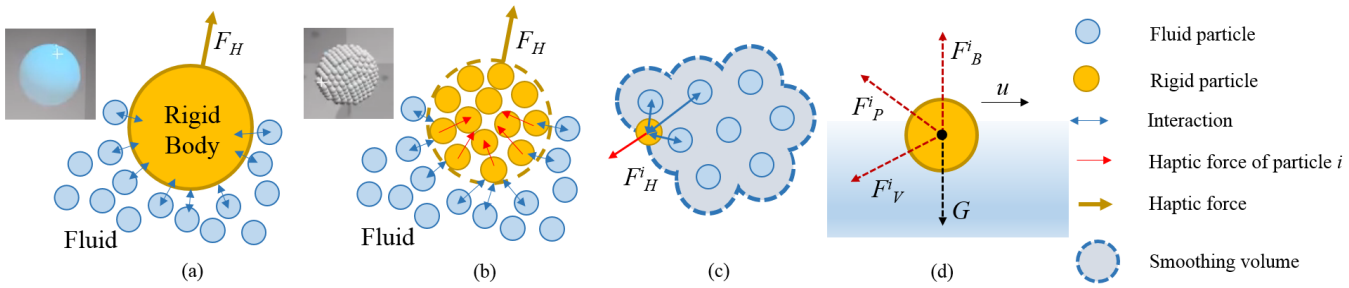


Figure 3: Illustrations of the SPH haptic rendering. (a) represents a rigid ball and volume of fluid interaction. (b) shows the rigid ball converted to a set of particles. The upper left corner is the real render frame in (a) and (b). (c) illustrates the smoothing volume of the volume of fluid. When a rigid particle enters the smoothing volume, the haptic force of particle i is calculated through the interaction (blue arrow) between SPH fluid and rigid body. (d) shows the forces on a rigid particle. The haptic force F^i_H is a combination of these different forces. Among them, the red arrows identify the forces we utilize in the sound synthesis algorithm.

$$\frac{d\rho}{dt} = -\rho \bigtriangledown u \tag{2}$$

where u denotes velocity of SPH particle, ρ represents density, P is pressure, F is the external force and μ is the viscosity coefficient. SPH fluid simulation is based on particles carrying different physical properties, such as mass and viscosity. The core of SPH method is that each particle has a smoothing kernel $W(\cdot)$ to determine the weight of the contributions of a neighbor particle. The smoothed quantity A at position x_i is approximated using a finite set of sampling points x_j located within a distance $x_{ij} = (x_i - x_j) < h$ as

$$A(x_i) = \sum_j m_j \frac{A_j}{\rho_j} W(x_{ij}, h) \tag{3}$$

where h is the smoothing radius, m_j is the mass of point j and ρ_j is the density of j. Readers can refer to [6, 7] for details. Since the quantity of SPH particles satisfy the mass conservation, SPH particle's acceleration a_i can be calculated by:

$$a_i = \frac{du_i}{dt} = \frac{F_i}{\rho_i} \tag{4}$$

where F_i is the sum of external forces on the right side of Equation (1), u_i is the velocity of particle i. With a_i, we can update position x_i and Equation (3).

Haptic force The haptic force comes from the interaction between SPH fluid and rigid body. Before we calculate the haptic force, we first classify the entities interacting with fluid. In physics-based fluid simulation, there are two types of entities, namely static entity (such as floors, walls), and dynamic entity (such as stick, movable ball). In our research, we need to handle both kinds of entities. For static rigid body, we assume it as static boundary and utilize boundary detection method to calculate resistance force [21]. Rigid body dynamics are more complicated than static rigid bodies.

To obtain highly effective, we solve the problem based on the unified particle model. As shown in Figures 3(a) and 3(b), we use the voxelization method to sample rigid body particles, and initialize property of rigid body. The motion of a dynamic rigid body can be seen as it moves and rotates around the centroid without deformation. Thus, we can employ a 6-DoF input haptic device to control rigid body in the virtual scene by changing the movement and rotation of rigid body into the rigid body particles' movement. Assuming that all the rigid particles share the same mass, the velocity of particle i can be expressed as:

$$u^i_r = u_r + \omega_r \times q_i \tag{5}$$

where u_r is the velocity of rigid body's centroid, ω_r represents the angular velocity of rigid body, u^i_r is the velocity of rigid body particle i, q_i denotes the relative position between particle i and the centroid. In each time step, we input the particle position and the rotation angle by haptic device to update position and velocity of rigid particles and fluid particles.

When the rigid body particles come into contact with the fluid particles (Figure 3(c)), that is, a rigid particle is inside the smoothing volume (defined as a sphere with the smoothing radius), we calculate the sum of haptic forces F_H based on the SPH simulation (Equation (8)). However, for a rigid particle i, the haptic force F^i_H is a combination of different forces which can be represented as $F^i_H = F^i_P + F^i_V + F^i_B$, where F^i_P, F^i_V and F^i_B represents pressure, viscosity and buoyancy on rigid particle i as illustrated in Figure 3(d). The calculation equations are as follows:

$$F^i_P = -\sum_j \frac{P_i + P_j}{2} \bigtriangledown W(x_{ij}, h) \tag{6}$$

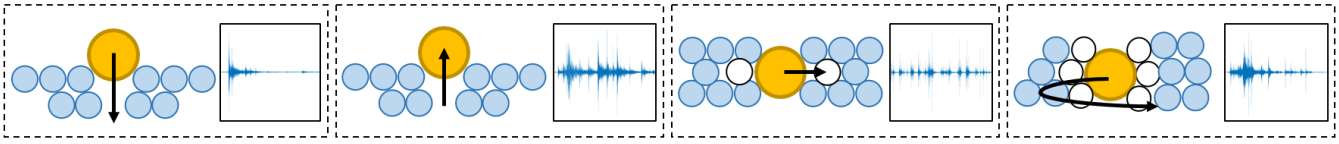
$$F^i_V = \mu V_i \sum_j V_j (u_i - u_j) \bigtriangledown^2 W(x_{ij}, h) \tag{7}$$

$$F^i_B = -\sum_j \rho_j d_j W(x_{ij}, h) \tag{8}$$

where V_i and V_j are the volume of i and j, d_j is the distance between the particle j and the liquid surface in the vertical direction, ρ_j represent the density of particle j. Readers can refer to [10, 19] for calculation details. We sum the forces on all rigid particles and get the pressure F_P, viscosity F_V and buoyancy F_B to guide the sound synthesis by force sequences.

4.2 MF granulation

The MF granulation is based on the haptic force which contains pressure F_P, viscosity F_V and buoyancy F_B generated in the previous subsection. Since the vibrotactile and acoustic phenomena share a common physical source, which makes it possible to guide the sound synthesis with haptic force. For the same physical phenomenon, although the feedback is different in different modalities, they are related to each other. For example, when we move the rigid ball in water faster, we feel stronger haptic force and also hear louder liquid sound. This is because increasing the relative velocity will increase the viscous force as shown in Equation (8). So although feedback forms of different modalities are different, the contents are relevant. Thus, in order to select the "correct" grain, we design a haptic force guided granulation algorithm to generate the grain in a controllable range. In this Section, we first analyze the mapping relationship between different forms of haptic and auditory perception for the same



(a) (b) (c) (d)

Figure 4: An illustration of grains. From (a) to (d): four types of grains, namely impact sound, leap out sound, bubble sound and cavity sound. In each sub-figure, the right part shows the waveform and the left part illustrates the corresponding motion of the rigid body. The yellow, blue and white circles represent rigid body particle, fluid particle and air particle, respectively. The black arrow marks the motion path of rigid body particle.

phenomenon. Then we design a novel MF granulation method which solved the inherent synchronization problem in granular synthesis.

Firstly, we classify the recordings, the standard of which is reference to the physical process of sound generation. The physical processes that generate sound during rigid-fluid interaction can be divided into three components [9]: (1) the initial high frequency impact, (2) the small bubble harmonics, and (3) the main cavity oscillation. However, in our experiments, we observe that the sound of rigid body leaping out of water can not be expressed by these three kinds of sound, so we expand the category to four categories as shown in Figure 4. Thus, the recordings include impact sound S_{impact}, leap out sound S_{leap}, bubble sound S_{bubble} for small bubble harmonics and cavity sound S_{cavity} for the main cavity oscillation.

To generate the grain in a controllable range, we design a set of controls to determine the position for extracting grains. The position Pos is classified into two parts, namely a fix start Pos_s and a random deviation $rand(\delta)$. Typically, the position Pos is defined as:

$$Pos = Pos_s + rand(\delta) \tag{9}$$

Then, we further analyze the force changes in these four situations to determine the Pos. The determinations are listed below:

- When $F_B^{t-1} = 0$ and $F_B^t > 0$, then $Pos_s \leftarrow S_{impact}(start)$.
- When $F_B^{t-1} > 0$ and $F_B^t = 0$, then $Pos_s \leftarrow S_{leap}(start)$.
- When $sum(F_B^0 : F_B^t) > 0$ and $|F_P^{t-1} + F_P^t| = |F_P^{t-1}| + |F_P^t|$, then $Pos_s \leftarrow S_{bubble}(start)$.
- When $sum((F_B^0 : F_B^t) > 0$ and $|F_P^{t-1} + F_P^t| \neq |F_P^{t-1}| + |F_P^t|$, then $Pos_s \leftarrow S_{cavity}(start)$.

Since the forces are calculated through position and velocity, each force sequence is a time series. Therefore, we express the force components at time t as F_*^t $(* = B, P, V)$. Here, we further explain the mapping scheme. At time t, when buoyancy appears, namely $F_B^{t-1} = 0$ and $F_B^t > 0$, the fix start Pos_s is set to be the start of the impact sound. When the buoyancy disappears, namely $F_B^{t-1} > 0$ and $F_B^t = 0$, the fix start Pos_s is set to be the start of the leap sound. When the fluid become wrinkled and the rigid body moves in the fluid in one direction, that is $sum(F_B^0 : F_B^t) > 0$ and $|F_P^{t-1} + F_P^t| = |F_P^{t-1}| + |F_P^t|$, the fix start Pos_s is set to be the start of the bubble sound. When the direction of pressure changes, that is, $|F_P^{t-1} + F_P^t| \neq |F_P^{t-1}| + |F_P^t|$, the fix start Pos_s is set to be the start of the cavity sound. The length of the deviation interval is equal to the grain size in four cases. Since t starts from 0, in order to ensure the stability of numerical computation, we assume $F_B^{-1} = 0$ and $F_P^{-1} = 0$. The linear mapping between force and Pos ensure the efficiency of the algorithm.

However, the sound of the rigid-fluid interaction varies with the physical property (i.e. density) of object and motion state (i.e. velocity). Therefore, we further adjust the amplitude of grains with $A^t = F_B^t \times F_V^t$. Among the four stages, buoyancy is mainly used to adjust the impact sound and leap out sound. Because buoyancy is invariable when an object enters the liquid completely, viscous force becomes the main influence factor when the object moves in the water.

Algorithm 1: Haptic force guided sound synthesis

Input: Four categories of recordings S_{impact}, S_{leap}, S_{bubble}, S_{cavity}, buoyancy F_B and pressure F_P.
Output: The generated $grain^t$ and final sound Out_S.
for *each time t* **do**
  Determine the Pos with F_B^t and F_P^t (Section 4.2);
  Extract the initial $grain^t$ from recordings S_{impact}, S_{leap}, S_{bubble}, S_{cavity};
  \\ Compute the spectrum of each grain ;
  $f_g \leftarrow \text{FFT}(grain^t)$;
  Phase reconstruction ;
  \\ Transform the spectrum to time-domain ;
  $grain^t \leftarrow \text{IFFT}(f_g)$;
  $Out_S^t \leftarrow grain^t$;
end

4.3 Spectral granular synthesis

With the generated grains, we can synthesize the final sound with the spectral granular synthesis method. Since in the VR application, the audio is real-time feedback with user operation, so it is impossible to hide artifacts through post-processing operations. Therefore, traditional time-domain granular synthesis methods which usually require a post-processing amplitude envelop for smoothing grains' edges are not suitable for our case. Facing this issue, we synthesize the sound in frequency domain with phase reconstruction [14] in the process of generation to ensure the continuity of the sound. Thus, in this section, we first determine the size of grains, then we utilize phase reconstruction for achieving continuity of the sound.

Grain size The size selection of grains is a tradeoff between the efficiency of algorithm and result quality, so we need to set appropriate size for the granular synthesis. Because the impact sound has obvious temporal structure, when we determine the size of grains, we mainly consider the impact sound duration and visual rendering. The frame rate for visual rendering is 25 frames per second and the audio rate is 48 kHz, that is, each frame corresponds to 1920 audio samples. A full impact sound is about 0.15 to 0.2 seconds, namely 7200 samples to 9600 samples. To ensure the integrity of S_{impact}, we use 9600 samples per grain, i.e., one grain corresponds to five frames.

Phase reconstruction To ensure the continuity of the sound result and provide more adjustment options, we reconstruct the phase in frequency domain. Compared to time-domain granular synthesis, with the intermediate step in the frequency domain, the phase can

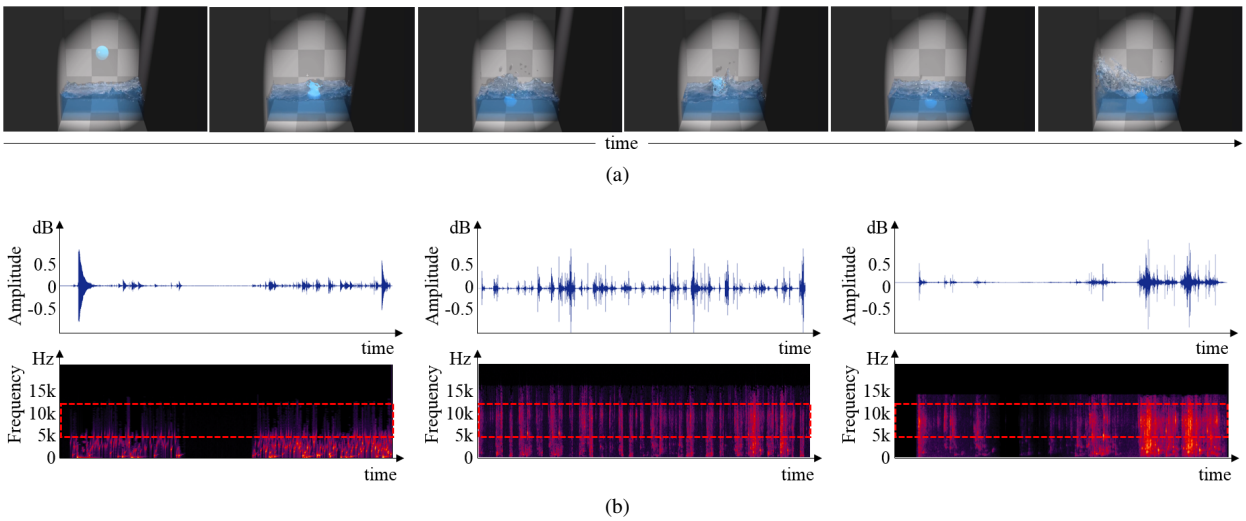


Figure 5: Comparison with the state-of-the-art methods. (a) presents six frames of the ball motion according to the time sequence. (b) illustrates the waveforms and spectra generated by (from left to right) the method of Cirio et al. [9], Fasciani [14] and our method. Red dotted boxes mark the different places in the spectra.

be reconstructed with compensation methods (i.e. Single Pass Spectrogram Inversion). The phase reconstruction step is equivalent to reconstructing a signal from a spectrogram, where phase information is not available. Readers can refer to [14] for details.

The whole process of the force-guided spectral granular synthesis is summarized in Algorithm 1. During the interaction, we play the sound $grain(t)$ in real time. At the end of the interaction, we can get the overlap-add final sound result Out_s. In the implement process, the sound examples are from a sound effects website [1]. Moreover, benefit from the reconstruction during MF granulation and spectral granular synthesis, we can generate interaction sounds for different velocities and densities of balls with one sample sound, which is evaluated in Section 5.2.

5 Results and Discussion

We tested our method on various rigid-fluid scenes. We implemented our algorithm on CUDA 7.0 and all the experiments were conducted with the same hardware: Intel Core i7-4790, NVIDIA GeForce GTX745 GPU, and 8GB RAM.

5.1 Comparison with the state-of-the-art methods

In order to validate our haptic force guided sound synthesis method, we compared it with the state-of-the-art methods, including physical based method [9] and the granular synthesis [14] in a multisensory simulation architecture. Figure 5 shows a comparison among synthesized sound results by [9], sound results by [14] and our method. The ball was put into water and then moving around in it. We presented five frames according to the time sequence as shown in Figure 5(a). The method in [9] is a physical based method with simplified bubble sound model which may reduce the authenticity of sound. From the waveform in Figure 5(b), we can observe that the waveform generated by the granular synthesis algorithm [14] is random which shows that the results are not matched the user interaction. Then we compared the timbre of different results, as shown in Figure 5(b) (bottom row), we can find that result generated by the method in [9] lacks high frequency information (see red dotted boxes). The difference in sound results can be heard in the accompanying video.

[1] https://www.audioblocks.com/

Table 1: Timing statistics for different methods (s)

Methods	Granulation	Phase reconstruction	Total
1:Fix [14]	0.001	0.041	0.067
2:Sequential [14]	0.003	0.034	0.058
3:Random [14]	0.001	0.033	0.065
4:Time-domain [16]	0.001	-	0.050
5:Ours	0.002	0.032	0.063

5.2 Evaluation

Algorithm efficiency Because the granular synthesis algorithm can achieve the highest computational efficiency among different kinds of sound synthesis algorithms, we compared our method with different state-of-the-art granular synthesis methods to evaluate that although we added haptic force to guide sound synthesis, the granulation process can achieve the real-time performance. Table 1 shows the time of generating 1-second sound by our method and other granular synthesis methods [14, 16] in our experimental environment. In the table, the top three rows are three different grain extraction methods (fix, sequential and random) in a spectral granular synthesis method [14], namely extracting grains in fix position, extracting grains in a sequential way and extracting grains randomly. The forth row presents the time of a time-domain granular synthesis method [16]. The right three columns represent the time for generating the grains, time for phase reconstruction (the time-domain method does not contain this step) and the total time. From the table we can observe that the total generation time is between 0.05-0.07 second, which ensures real-time feedback. Thus, compared with the traditional granular synthesis method, our method can not only generate ideal sound sequence, but also ensure the real-time performance.

Different rigid-fluid interaction scenes In Figure 6, we presented three results to evaluate our haptic force guided sound synthesis method in a multisensory simulation system. In the lower left corner of each frame, we showed the image of the user interaction with haptic device. The three scenarios correspond to the movement of rigid ball in water, the falling of three balls of different densities

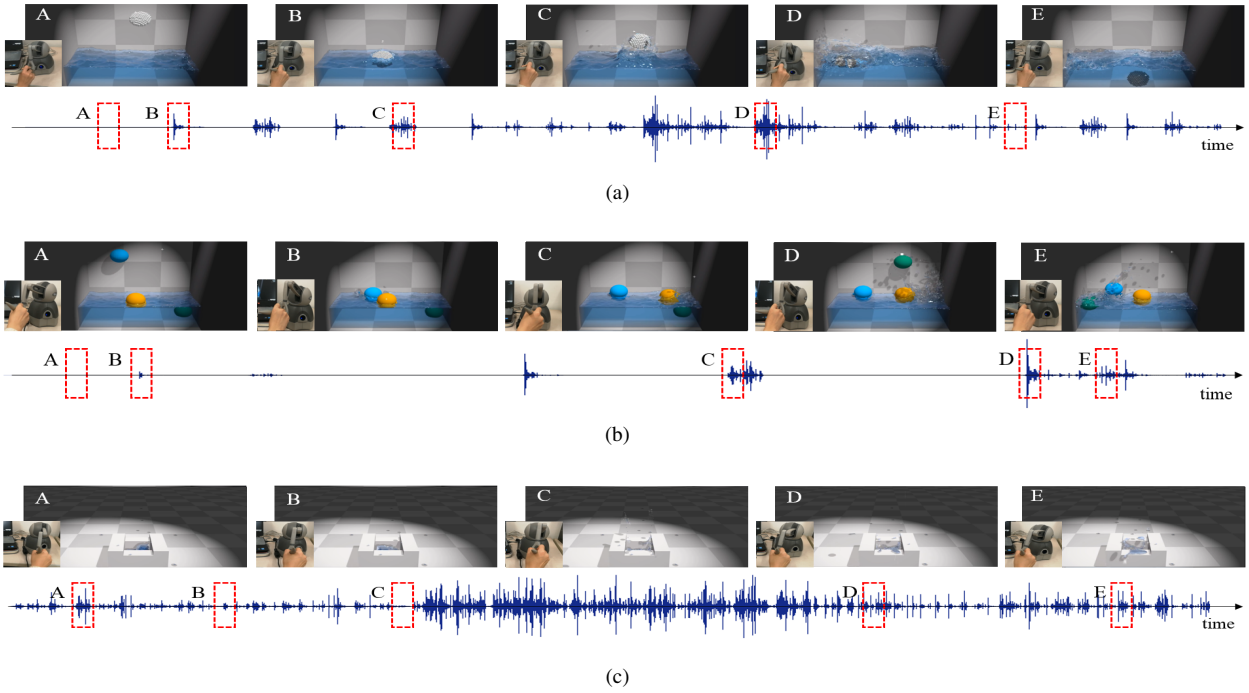


Figure 6: Three results for a multisensory simulation system that integrates our sound synthesis algorithm. In each sub-figure, the top row presents five frames of fluid animation. The bottom left corner inset in each frame shows the corresponding haptic scene. The bottom row illustrates the waveform of the generated sound result and the red dotted boxes mark the generated audio synchronized with the visual animation.

and the movement of a rigid ball obscured by a container. From the waveforms in Figure 6, we can see that our synthesized sound results can accurately synchronize with the movement of rigid body, which greatly enhances the user immersion. More concretely, sound can help to determine some invisible information, such as density or velocity. As presented in Figure 6(b), when the three rigid balls have the same volume but different densities, the rigid-fluid interaction sound can help us to distinguish which ball has higher density. In Figure 6(c), when the rigid ball is occluded by the container, we can judge the movement information of the ball through the interaction sound. Moreover, all the results shown in Figure 6 used a same recording, but the generated sound showed that we could avoid the simple repetition with MF granulation and phase reconstruction.

5.3 Subjective Measures

To further assess the effectiveness of our approach, we designed three subjective measures to evaluate the quality of the synthesized sound. A total of 15 participants (7 female; 8 male) are assigned to the experiment. Participants are either undergraduate or graduate students and the average age of participants was 23.5 (SD = 1.3). All participants have either normal or corrected-to-normal vision and reported normal hearing ability.

Comparison with the state-of-the-art method First, in order to compare the reality of the synthesized sound results by our method and physical based method [9], we further designed a subjective experiment. Since we focus on comparing the quality of sound, in order to ensure the consistency of visual feedback, we adopted an off-line comparison scheme. We preserved the SPH particle motion information, buoyancy and pressure during the interaction, and then calculate the corresponding sound according to the algorithm in [9] and our work. In this way, the sound is the single variable in the experiment. Every participant was presented 5 pairs of videos, each pair of videos has different sounds that are generated by method in [9] and our method, respectively. In both cases, the participant was asked " Are these two video clips the same or different?" If the answer was "different", then we asked " Which video clip do you prefer?" and "What is the strength value of the preference?" Table 2 shows the results of the experiment. The score of strength value is on a scale from 1 to 10, where 1 is labeled "A little prefer" and 10 "Very prefer". From the results, as indicated by the " mean strength", we can observed that the participants clearly prefer our method.

In addition, a paired T-test is performed on these scores to check if the score of the synthesized sound by our method is significantly bigger than the score of the synthesized sound by [9] using the following hypotheses $H_0 : \mu_a \leq \mu_b$, $H_1 : \mu_a > \mu_b$, where μ_a represents the score of the synthesized sound by our method, while μ_b represents the score of the synthesized sound by [9]. Hypothesis H_0 means that results by [9] have a higher score. Table 2 shows all 5 paired T-test results. Note that all P values are less than 0.0005, and all T values are higher than 1.6839, indicating that H_0 is rejected with statistical significance while H_1 is accepted. This concludes that our result achieves significantly better sound effect than [9].

Evaluation of the user immersion Second, we designed an experiment to measure the effect of adding audio feedback on user immersion. In this experiment, the participant was shown a series of scenarios with and without audio feedback and the scenarios are listed in Table 2. In the experiment, participants were asked to answer the following questions:

- Can you recognize that you are interacting with the water, despite the rigid lever provides conflicting perceptual cues?
- Do you think increasing haptic feedback / haptic and audio feedback can increase reality?
- How strong is the sense of reality?

For the first two questions, we used 1 for "yes" and 0 for "no". For the third question, the score of strength value is on a scale from 1 to

Table 2: Subjective results for Experiment 1: sound synthesized by our method vs. sound synthesized by [9]

Scenarios	Prefer Ours	Prefer Cirio et al. [9]	Mean Strength of Our Method	Mean Strength of Cirio et al.'s method [9]	T Value	P Value
Falling ball	73.3%	26.7%	6.45	2.00	6.231	*****
Fast move	93.3%	6.7%	7.57	2.50	5.296	*****
Slow move	80.0%	20.0%	7.08	2.00	5.931	*****
Leaping out	73.3%	26.7%	7.54	3.31	8.570	*****
Hybrid	93.3%	6.7%	7.47	2.21	8.038	*****

Significance codes (P): < 0.05: '***', < 0.01: '****', < 0.005: '*****'

Table 3: Subjective results for Experiment 2: with audio vs. without audio

Scenarios	Mean (Q1)	Std (Q1)	Mean (Q2)	Std (Q2)	Mean (Q3)	Std (Q3)	P Value	T Value
Falling ball	0.73	0.2095	0.67	0.2381	6.33	2.5238	-	-
Falling ball (No audio)	0.33	0.2381	0.40	0.2571	3.47	1.4095	6.431	*****
Fast move	0.80	0.1714	0.73	0.2095	7.21	2.7642	-	-
Fast move (No audio)	0.27	0.2095	0.33	0.2381	4.23	1.2134	5.164	*****
Slow move	0.87	0.1238	0.87	0.1238	7.52	2.7112	-	-
Slow move (No audio)	0.33	0.2381	0.27	0.2095	3.44	1.3314	6.251	*****
Density	0.67	0.2381	0.73	0.2095	6.95	2.6629	-	-
Density (No audio)	0.20	0.1714	0.20	0.1714	3.27	1.2933	7.312	*****
Boat	0.73	0.2095	0.87	0.1238	6.31	2.5141	-	-
Boat (No audio)	0.27	0.2095	0.33	0.2381	2.67	1.1926	5.947	*****

Significance codes (P): < 0.05: '***', < 0.01: '****', < 0.005: '*****'

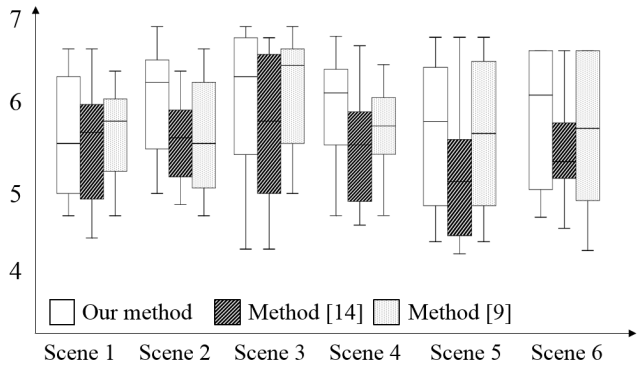


Figure 7: Subjective results for Experiment 3. Boxplots show for each scene the median, the interquartile range, minimum and maximum values.

10, where 1 is labeled "Not reality" and 10 "Very reality". Results are also analyzed using standard tests well suited for each study, namely a one-way Analysis of Variance (ANOVA) test for experiment 2, and are reported in Table 3. In both cases, statistical significance can be claimed if the reported p-value is less than a significance level of 0.05. For Post-hoc analysis (in our case, pairwise comparisons between conditions), a standard approach is to use the Bonferroni method and Wilcoxon signed-rank tests respectively, with a significance level adjusted for multiple comparisons.

Evaluation of the synchronization The third experiment is used to verify the synchronization of our sound results with visual and haptic feedback. To verify the effectiveness of our algorithm, there are six scenes in the experiment. In different scenarios, the motion of rigid body is different. We used the random mapping algorithm in [14], physical based method [9] and our force-guided sound synthesis method in different scenes. The participants are asked to rate the synchronization on a scale from 1 to 10, where 1 is labelled as "Not synchronized" and 10 "Very synchronous." Figure 7 shows the score of the experiment. It can be observed that the score of our method is obviously higher than that of the random algorithm and similar to the results of physical methods, which means that the sound results generated through our method can effectively synchronize with visual and haptic feedback.

6 Conclusion and Future Works

In this work, we proposed a real-time haptic force guided sound synthesis method which is adequate for multisensory VR simulation. Our method proved that results of different modal feedback can interact with each other. This provided us a new way to design an interactive approach in multisensory scenarios. We can obtain the state of another modality through a modal change. Since the analysis of motion directly from objects is complex, the indirect solution we proposed can make the synergy between different modalities more efficient. Through the novel haptic force guided sound synthesis method, we could achieve higher quality sound results with higher time performance than the state-of-the-art sound synthesis solutions for multisensory VR systems. We designed different example scenarios to evaluate the interaction possibilities offered by our method. The results proved the effectiveness of our algorithm.

In addition, by changing the sample sounds and force types, our method could also be used for sound synthesis of rigid-rigid interaction scenes and rigid-deformable interaction scenes. But this also shows the limitation of our approach, i.e., the generated sounds would be limited by the recordings. Hence, combining physics-based modelling and haptic force guided sound synthesis is a research direction of our future work. We would calculate different sounds through physical model in pre-processing. Moreover, the relationship between different modalities also deserves further exploration.

Acknowledgments

This work was supported by the Natural Science Foundation of China under grant nos. 61170118 and 61672375, the National Key R&D Program of China under no. 2018YFC1407405, and the Open Project Program of State Key Laboratory of Virtual Reality Technology and Systems, Beihang University under no. VRLAB2018A15.

REFERENCES

[1] T. Amada, M. Imura, Y. Yasumuro, Y. Manabe, and K. Chihara. Particle-based fluid simulation on GPU. In *ACM Workshop on General-Purpose Computing on Graphics Processors*, pp. 342–343, 2004.

[2] R. Ando, N. Threy, and R. Tsuruno. Preserving fluid sheets with adaptively sampled anisotropic particles. *IEEE Transactions on Visualization and Computer Graphics*, 18(8):1202–1214, 2012.

[3] F. Avanzini and P. Crosato. Integrating physically based sound models in a multimodal rendering architecture. *Computer Animation and Virtual Worlds*, 17(3-4):411–419, 2006.

[4] M. Becker, H. Tessendorf, and M. Teschner. Direct forcing for lagrangian rigid-fluid coupling. *IEEE Transactions on Visualization and Computer Graphics*, 15(3):493–503, 2009.

[5] A. Bloomfield, Y. Deng, J. Wampler, P. Rondot, D. Harth, M. Mcmanus, and N. Badler. A taxonomy and comparison of haptic actions for disassembly tasks. In *IEEE Virtual Reality*, pp. 225–231, 2003.

[6] R. Bridson. *Fluid simulation for computer graphics*. A K Peters/CRC Press, 2008.

[7] D. Charypar and M. Gross. Particle-based fluid simulation for interactive applications. In *ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pp. 154–159, 2003.

[8] G. Cirio, M. Marchal, S. Hillaire, and A. Lcuyer. Six degrees-of-freedom haptic interaction with fluids. *IEEE Transactions on Visualization and Computer Graphics*, 17(11):1714–1727, 2011.

[9] G. Cirio, M. Marchal, A. Lecuyer, and J. R. Cooperstock. Vibrotactile rendering of splashing fluids. *IEEE Transactions on Haptics*, 6(1):117–122, 2013.

[10] G. Cirio, M. Marchal, M. A. Otaduy, and A. Lcuyer. Six-Dof haptic interaction with fluids, solids, and their transitions. In *World Haptics Conference*, pp. 157–162, 2013.

[11] G. D. Theory of communication. part 1: The analysis of information. *Journal of the Institution of Electrical Engineers-Part III: Radio and Communication Engineering*, 93(26):429–441, 1946.

[12] Y. Dobashi, M. Sato, S. Hasegawa, T. Yamamoto, M. Kato, and T. Nishita. A fluid resistance map method for real-time haptic interaction with fluids. In *ACM Symposium Virtual Reality Software and Technology*, pp. 91–99, 2006.

[13] K. V. D. Doel. Physically based models for liquid sounds. *ACM Transactions on Applied Perception*, 2(4):534–546, 2005.

[14] S. Fasciani. Spectral granular synthesis. In *International Computer Music Conference*, pp. 1–5, 2018.

[15] F. Florian, A. Ryoichi, W. Chris, W. Rdiger, and T. Nils. Narrow band FLIP for liquid simulations. *Computer Graphics Forum*, 35(2):225–232, 2016.

[16] M. Frjd, , and A. Horner. Sound texture synthesis using an overlapadd/granular synthesis approach. *Journal of the Audio Engineering Society*, 57(1/2):29–37, 2009.

[17] M. S. Howe. *Acoustics of fluid-structure interactions*. Cambridge University Press, 1998.

[18] T. R. Langlois, C. Zheng, and D. L. James. Toward animating water with complex acoustic bubbles. *ACM Transactions on Graphics*, 35(4):95:1–95:13, 2016.

[19] S. Liu, C. Ma, and G. Feng. Haptic rendering for the coupling between fluid and deformable object. *Virtual Reality*, 22(3):1–12, 2018.

[20] N. Ma, Y. Liu, A. Qiao, and J. Du. Design of three-dimensional interactive visualization system based on force feedback device. In *International Conference on Bioinformatics and Biomedical Engineering*, pp. 1780–1783, 2008.

[21] M. Moore and J. Wilhelms. Collision detection and response for computer animation. *Computer Graphics*, 22(4):289–298, 1988.

[22] J. Mora and W. S. Lee. Real-time fluid interaction with a haptic device. In *IEEE International Workshop on Haptic, Audio and Visual Environments and Games*, pp. 160–165, 2007.

[23] J. Mora and W. S. Lee. Real-time 3D fluid interaction with a haptic user interface. In *IEEE Symposium on 3D User Interfaces*, pp. 75–81, 2008.

[24] W. Moss, H. Yeh, J. M. Hong, M. C. Lin, and D. Manocha. Sounding liquids: automatic sound synthesis from fluid simulation. *ACM Transactions on Graphics*, 29(3):21:1–21:13, 2010.

[25] N. C. Nilsson, R. Nordahl, L. Turchet, and S. Serafin. Audio-haptic simulation of walking on virtual ground surfaces to enhance realism. In *International Conference on Haptic and Audio Interaction Design*, pp. 61–70, 2012.

[26] R. Nordahl, S. Serafin, and L. Turchet. Sound synthesis and evaluation of interactive footsteps for virtual reality applications. In *IEEE Virtual Reality*, pp. 147–153, 2010.

[27] S. Papetti, F. Fontana, M. Civolani, A. Berrezag, and V. Hayward. Audio-tactile display of ground properties using interactive shoes. In *International Conference on Haptic and Audio Interaction Design*, pp. 117–128, 2010.

[28] E. G. Richardson. The sounds of impact of a solid on a liquid surface. *Proceedings of the Physical Society*, 61(4):541–547, 1955.

[29] M. Sagardia and T. Hulin. Evaluation of a penalty and a constraint-based haptic rendering algorithm with different haptic interfaces and stiffness values. In *IEEE Virtual Reality*, pp. 64–73, 2017.

[30] M. Sagardia, B. Weber, T. Hulin, G. Hirzinger, and C. Preusche. Evaluation of visual and force feedback in virtual assembly verifications. In *IEEE Virtual Reality*, pp. 23–26, 2012.

[31] B. Solenthaler, J. Schlfli, and R. Pajarola. A unified particle model for fluid-solid interactions. *Computer Animation and Virtual Worlds*, 18(1):69–82, 2010.

[32] B. Truax. Real-time granular synthesis with a digital signal processor. *Computer Music Journal*, 12(2):14–26, 1988.

[33] B. Truax. Music and science meet at the micro level: Time-frequency methods and granular synthesis. *Acoustical Society of America Journal*, 117(4):2415–2416, 2005.

[34] C. Verron and G. Drettakis. Procedural audio modeling for particle-based environmental effects. In *133rd AES Convention*, pp. 1–18, 2012.

[35] K. Wang, H. Cheng, and S. Liu. Efficient sound synthesis for natural scenes. In *IEEE Virtual Reality*, pp. 303–304, 2017.

[36] K. Wang and S. Liu. Example-based synthesis for sound of ocean waves caused by bubble dynamics. *Computer Animation and Virtual Worlds*, 29(4):e1835, 2018.

[37] P. Xia. New advances for haptic rendering: state of the art. *Visual Computer*, 34(3):1–17, 2016.

[38] M. Yang, J. Lu, A. Safonova, and K. J. Kuchenbecker. GPU methods for real-time haptic interaction with 3D fluids. In *IEEE International Workshop on Haptic Audio Visual Environments and Games*, pp. 24–29, 2009.

[39] X. Zhang and S. Liu. SPH haptic interaction with multiple-fluid simulation. *Virtual Reality*, 21(4):1–11, 2017.

[40] C. Zheng and D. L. James. Harmonic fluids. *ACM Transactions on Graphics*, 28(3):37:1–37:12, 2009.