CHAPTER 2: BACKGROUND

This chapter briefly discusses the background knowledge required to understand the work presented in subsequent chapters.

2.1 Sound Waves and Harmonic Motion

We first discuss sound waves and harmonic motion. When an object is struck, it vibrates and deforms. The surrounding air rapidly compresses (compression) and expands (rarefaction or decompression) as the object vibrates outward and inward respectively, As it oscillates periodically, pressure waves are created and air pressure amplitude changes up and down over time. This periodic pattern of compression and rarefaction is known as harmonic motion. Although we may not see the vibrations or deformations, our ears hear the variation in air pressure as sound.

2.1.1 Spring-Mass System

Like harmonic motion, a spring mass system also responds in a period motion. For the underdamped case ($0 < d^2 < 4 * k * m$) where d is damping, k is stiffness, and m is mass, sound amplitude vibrates but eventually stops. Equation 1.1 models the displacement for a single mass. For a volumetric object, the system can be modeled by the sound dynamics equation:

$$Mu'' + Du' + Ku = f \tag{2.1}$$

where M is the mass matrix, where mass is located on the object, u is the displacement of each element, D is the viscous damping matrix (i.e. how velocity decays over time), K is the stiffness matrix (i.e. defining the connectivity of the elements) and f is the vector of forces (i.e. inducing vibrations). Note that upper case denotes matrices, lower case denotes vectors or scalars, and M, D, and K are size 3n x 3n sparse matrices for n tetrahedral mesh nodes. Stiffness is based on the objects mesh and Poisson's ratio; mass, construction method Consistent Mass Matrix (CMM); and damping, Rayleigh damping (also known as linearly proportional damping).

$$D = \alpha_1 * M + \alpha_2 * K \tag{2.2}$$

where α_1 and $alpha_2$ are real-value parameters. Given these parameters, we can simulate the vibration of the solid volume body object in response to an impulse.

2.1.2 Modal Analysis and Sound Synthesis

To achieve real-time performance, modal analysis performs pre-processing steps for a given object and material. Then at runtime, sounds are dynamically created with modal synthesis based on hit (or contact) point where the object is struck and impulse direction. Typically, we simulate an impulse direction normal to the contact point but could synthesize tangentially to the object like (0,1,0) for (x,y,z) impulse direction. The solution to the sound dynamics equation is damped sinusoidal waves.

$$q_i = a_i * e^{-d_i * t} sin(2\pi * f_i * t + \theta_i) \tag{2.3}$$

where q_i is the displacement, a_i depends on run-time impulse, and d_i along with f_i depend on geometry and material properties.

$$d_i = \frac{1}{2}(\alpha_1 + \alpha_2 * \lambda_i) \tag{2.4}$$

$$f_i = \frac{1}{2\pi}\sqrt{\lambda_i - (\frac{\alpha_1 + \alpha_2 * \lambda_i}{2})^2} \tag{2.5}$$

2.2 Fluid-Structure Interaction and Elasto-Acoustic Coupling

Previous research focuses on single systems only, either solid or liquid but not both. Elasto-acoustic coupling refers to a solid-fluid interface. It involves the interactions between the vibrations of an elastic structure and the sound field in the surrounding fluid. Using an added mass operator simplifies a coupled problem into a single fluid-structure system.

The added mass operator assumes a non-moving domain; that is, the structural vibration must move the liquid along with the structure. Then, the weight of the surrounding liquid can be added to the system by modifying the mass matrix of the structure object. This is referred to as a rigid double body where the added mass is the additional (drag) force resulting from fluid acting on a structure. Since the liquid In reality, the fluid will be accelerated but for simplicity, liquid is modeled as a volume moving with the object.

$$Mu'' + Du' + Ku = f(t) - m_a * u'' \tag{2.6}$$

2.3 Multimodal Learning with Audio Visuals

Multimodal learning explores relationships between different modalities. These techniques are found in natural language processing methods such as visual question and answering systems that use both images and text. Recently, audio-visual learning has also been used for speech separation by using both the image of the speakers and their speech. In addition to language, multimodal learning from impact sounds of meshed geometries have also been demonstrated to aid in localization, object, and material classification.

Learning based on multiple modes also enables methods to fill missing modalities. For example, while vision-based methods may have difficulty detecting or estimating depth for textureless or glass surfaces, echoes of reflecting sounds from a sound source may fill those holes. Furthermore, reconstruction methods may also benefit from audio for enhanced geometry reconstruction and material inference.

2.4 3D Reconstruction Systems

A common processing pipeline involves capture, point generation, meshing and texturing, and temporal mesh processing. During capture, the setup often includes some level of calibration, preprocessing, bias correction, and background subtraction. 3D point cloud generation follows from depth maps (IR, RGB, SfS), multimodal MVS, or iterative surface estimation (point-based). Meshing and texturing may include topology denoising, island removal, hull-constrained PSR, occlusion detection, and texturing. Finally, temporal mesh processing incorporates mesh tracking, importance detection, and mesh decimation.