Feature Based Impact Sound Synthesis of Rigid Bodies Using Linear Modal Analysis for Virtual Reality Applications

Article *in* Journal of the Audio Engineering Society · February 2016

Muhammad Imran
RWTH Aachen University

Jin Yong Jeon
Hanyang University

FEATURE BASED IMPACT SOUND SYNTHESIS OF RIGID BODIES USING LINEAR MODAL ANALYSIS FOR VIRTUAL REALITY APPLICATIONS

MUHAMMAD IMRAN[1], **JIN YONG JEON**[1]

[1] *Department of Architectural Engineering, Hanyang University, Seoul, Korea*
`mimranh1@gmail.com, jyjeon@hanyang.ac.kr`

This paper investigates an approach for synthesizing sounds of rigid body interactions using linear modal synthesis (LMS). We propose a technique based on feature extraction from one recorded audio clip to estimate perceptually satisfactory material parameters of virtual objects for real-time sound rendering. In this study, the significant features from one recorded audio are extracted by computing high level power spectrogram that is based on short time Fourier transform analysis with optimal windowed function. Based on these reference features, the material parameters of intrinsic quality are computed for interactive virtual objects in graphical environments. A tetrahedralize finite element method (FEM) is employed to achieve Eigen values decomposition during modal analysis process. Residual compensation is also implemented to optimize the perceptual differences between the synthesized and the real sounds, and to include the non-harmonic components in the synthesized audio in order to achieve perceptually high quality sound. Furthermore, the computed parameters for material objects of one geometry can be transferred to different geometries and shapes of the same material objects, whereas, the synthesized sound varies as the shapes of the objects change. The results of the estimated parameters as well as a comparison of real sound and the synthesized sound are presented. The potential applications of our methodology are synthesis of real time contact sound events for games and interactive virtual graphical animations and providing extended authoring capabilities.

INTRODUCTION

Virtual sound synthesis models for graphical objects have played a significant role in real-time sound production for different types of interactions among the rigid bodies. Recent developments have been made in automatic virtual sound synthesis for collision, rolling and frictional interactions. However, to achieve a good quality audio in animated simulations and for interactive video games it remains a challenging task to represent the real-world complements during the synthesis processes. The basic challenge is in emulating real world sounds accurately with the synthesized sounds because of inappropriate information of real sound effects. Moreover, there are certain discrepancies in synchronization of audio and video during rendering process that create disparity among different audio and visual cues and hence produces a poor virtual perception. As a results, audio-video perceptual cues lose its plausibility. The traditional approaches relay on manual recording and editing the sound effects to a visual scene prior to the synchronization with video games, animated scenarios and movies. These approaches produce adequately satisfactory results, however, these are not applicable to interactive environments in real-time applications.

Recently, granular synthesis techniques have been widely adopted as sound synthesizers with computers in which tiny grains of sounds are created to create sound signals that resemble a particular event [1]. Gabriel in [2] investigated an approach of dynamic Level of Audio Detail (LOAD), where, a complex sound scene can be synthesized with much less computation and storage than would be required to deliver, process and render a large number of samples.

Alternatively, physically-based sound synthesis methods have been proposed by many researchers to create a realistic sound in synchronization with the visual cures. Different interesting sounds of fluid dynamics have been suggested by Zheng et. al. [3]. Picard et al. [4] proposed a method for auto-analysis of audio clips, and the generation of compact dictionaries of audio grains and correlation patterns in a physics engine. A rigid body interaction approach is proposed by O'Brien [5] that simulates rigid body vibrations which leads to variations in perceptual sound pressures. In this approach accurate capturing of surface vibrations and wave propagations are measured. This approach is inefficient to handle interactive real-time applications. For run-time analysis, linear modal synthesis (LMS) has been broadly recognized to synthesize sounds for rigid-body interactions, where, a modal model is required in order to generate a sound at real-time excitations. These models provide different complex interaction handlings (e.g., impact, rolling and sliding) to create highly realistic sounds. However, the earlier virtual sound synthesis research focused on estimating material parameters for modal analysis in order to recreate realistic sounds.

In this paper an efficient sound synthesis technique is applied that produces the real and high quality virtual sound by manipulating the physically based modal synthesis process. In this approach, an impact sound is recorded as a specimen audio and we extract the significant features in terms of frequency, damping and initial amplitude of the dominant damped harmonics from this audio clip. These features are used during the process of estimation of the significant material parameters of the graphical objects of the same size and shape as that of real world objects. Once the material parameters (e.g., stiffness of material, damping coefficients, and mass density) are estimated several modes are generated for different interactions of bodies with various geometries in runtime.

1 MODAL ANALYSIS

For real time applications in games, much work has been done on modal synthesis for sound rendering processes. The modal synthesis processes are independent of any recorded sound to create virtual sound that is generated by interactions among graphical objects of a game. In this way, modal analysis does not require manual synchronization of the audio-visual events and the synthesized sound reflects a compromising variations of interactions among game objects of different geometries. O'Brien in [5] proposed a framework for rigid body impact sound handling in which the graphical virtual objects are represented as tetrahedral mesh geometries, therefore, a tetrahedral finite element model (FEM) is applied to a given object of any shape. The displacement vectors are computed with the linear deformation equation, given by Equation 1.

$$\boldsymbol{M}\ddot{x} + \boldsymbol{C}\dot{x} + \boldsymbol{K}x = f \tag{1}$$

Here $\boldsymbol{M}$, $\boldsymbol{C}$, and $\boldsymbol{K}$ are the mass, damping and stiffness matrices respectively. As the surface vibrations during the contact of virtual bodies produce small scale level damping, therefore, an approximation to the damping matrix is considered by Rayleigh approximation. In other words damping matrix can be represented as a linear combination of mass matrix ($\boldsymbol{M}$) and stiffness matrix ($\boldsymbol{K}$), given by Equation 2.

$$\boldsymbol{C} = \alpha\boldsymbol{M} + \beta\boldsymbol{K} \tag{2}$$

The generalized eigenvalue problem, given in Equation 3, is then solved to decouple this system into the form described in Equation 4.

$$\boldsymbol{KU} = \Lambda\boldsymbol{MU} \tag{3}$$

$$\ddot{q} + (\alpha I + \beta\Lambda)q + \Lambda\dot{q} = \mathbf{U}^T f \tag{4}$$

Here, '$\boldsymbol{\Lambda}$' is a matrix with the eigenvalues of Equation 3, and '$\mathbf{U}$' is the eigenvector matrix that converts 'x' to the decoupled deformation bases 'q' and represented as '$x = Uq$'. The solutions to the Equation 4 are damped sinusoidal waves. The ith mode is given in Equation 5.

$$q_i = a_i e^{-d_i t}\sin(2\pi f_i t + \theta_i) \tag{5}$$

Here, 'f_i' represents the frequency and 'd_i' represents the damping coefficients of the ith mode, whereas, 'a_i' is the amplitude, and 'θ_i' is the initial phase. In Equation 5 the frequencies, damping and amplitudes are referred as the feature 'φ_i' of ith mode and represented as,

$$\varphi_i = (f_i, d_i, a_i) \tag{6}$$

These features, as expressed in Equation 6, depend on the material properties, geometries and the exciting forces during the real-time interactions. The general formulation to compute these features is given below in Equation 7 and Equation 8.

$$d_i = \frac{(\alpha + \beta\lambda_i)}{2} \tag{7}$$

$$\varphi_i = \frac{1}{2\pi}\sqrt{\lambda_i - \left(\frac{\alpha + \beta\lambda_i}{2}\right)^2} \tag{8}$$

2 METHODOLOGY

The basic structure of the methodology is explained in Figure 1, where, we recorded one sound clip for one kind of material and then estimated material parameters for that clip which are used during modal synthesis process in order to generate a sound. The synthesized sound reflects the same sense of perception as original recorded sound [6]. As a first step, once a sound is recorded for a certain real material, it is subjected to the feature extraction process for computing reference features (i.e., damped sinusoids of different frequencies, damping coefficients and initial amplitudes). Secondly, a virtual object of same size and shape, as that of real one, is created in physics engine and is decomposed into tetrahedral mesh geometry. Modal synthesis process, describe in section 2 is applied and the reference features are utilized in estimating the material parameters for this virtual object. For every impact of the virtual object, the estimated material parameters are used to generate the virtual sound. Finally, the difference metrics are calculated by a comparison of perceptual similarity between measured sound and virtually generated audio of objects with the residual compensation process. As an implication these material parameters can be transferred to different objects of varying geometries and shapes. The whole process is illustrated in Figure 1.

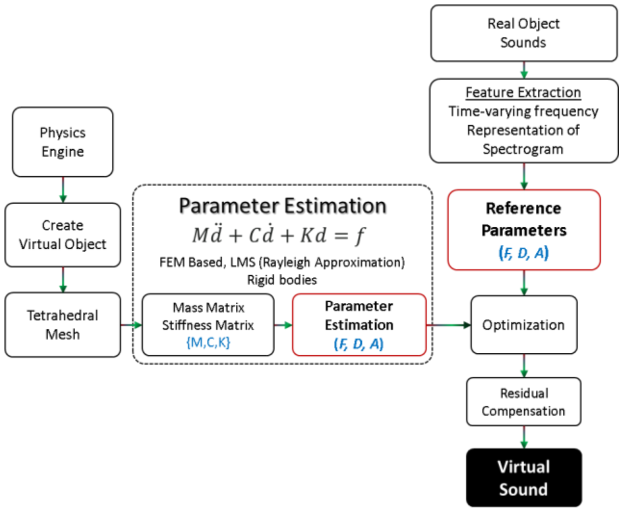

Figure 1: Computational Process and Methodology

2.1 Computational framework

This section describes the computational framework of proposed methodology. First, the audio clips for different rigid-bodies impact sound are recorded in an anechoic chamber hall and subsequently subjected to the feature extraction processing chain. High level features are extracted with spectrogram analysis of these sound clips by splitting them into short intervals of overlapping window frames in time domain. A power spectrogram is obtained by applying short time Fourier transform (STFT) using 'Hann' window fune41ction on each frame of the sound clip. We applied a searching algorithm to find out potential peaks in the time-frequency spectrogram. These peaks correspond to the potential modes of the recorded sound in the form of damped sinusoids. These potential modes are filtered out based on the criteria of selection the strongest peaks for a specified time window. The features are detected in the form of frequencies, damping, and amplitude envelopments '$\varphi_i = (f_i, d_i, a_i)$' known as reference features.

Secondly, a material parameter estimation algorithm is applied based on optimization process. These parameters are represented as '$\underline{\boldsymbol{\varphi}}_i = (\underline{\boldsymbol{f}}_i, \underline{\boldsymbol{d}}_i, \underline{\boldsymbol{a}}_i)$'. To calculate these parameters, we created a virtual graphical object of same size and geometry as that of real object for which the features were extracted. This virtual object is tetrahedralized to compute the mass matrix ($\boldsymbol{M}$) and stiffness matrix ($\boldsymbol{K}$) by providing initial values of the Young's modulus 'E_o', Poisson's ratio 'υ_o', and mass density 'ρ_o'. The eigenvalues 'λ_i' are computed as follows.

$$\lambda_i = \frac{\gamma}{\gamma_0}\lambda_i^0 \qquad (9)$$

Here $\gamma = E/\rho$ is Young Modulus ratio and $\gamma_0 = E_0/\rho_0$ is the ratio of initially assumed values to mass and density. A unit impulse, denote by 'a_o^j', is applied on the object that generates excitations of the eigenvalues as given in Equation 5. Finally, the parameters are computed by combining the Equations 7, 8 and 9. The sound is synthesized using the estimated parameters from Equation 10 given below.

$$\boldsymbol{s[n]} = \sum_{\boldsymbol{j}} \left(\boldsymbol{a_i e^{-d_i\left(\frac{n}{F_s}\right)} sin}\left(\boldsymbol{2\pi f_i}\left(\frac{\boldsymbol{n}}{\boldsymbol{F_s}}\right)\right)\right) \qquad (10)$$

The advantage of this approach is that the properties of recorded audio can be transferred to different objects of different sizes and of various geometries.

In the final step residual compensation is applied in order to compensate the non-harmonic part present in the real sound example clip. Linear modal approach is not capable for detecting the non-harmonic sound parts during the synthesis process, therefore, the quality of synthesized sound remains un-natural. Lloyd in [7] proposed a method for residual compensation by tracking the modal part in original sound and simply subtracting the power spectrograms of real and synthesized sounds. In our method, we followed the technique proposed by [8] for residual compensation based on subtracting the representative sound, synthesised from material parameters, from the original sound.

3 RESULTS

3.1 Features extraction

Figure 2 shows the spectrogram of recorded audio example of a glass impact sound whereas Figure 3 shows the tracks of the detected potential modes from STFT analysis using Hann window function.

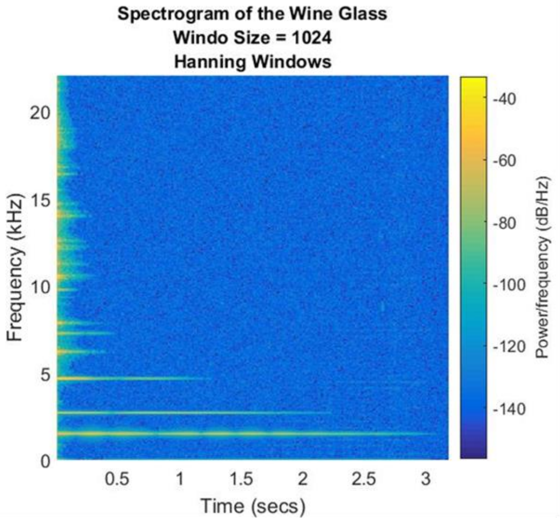

Figure 2: Example audio clip spectrogram

The threshold value for detecting the peaks of the potential modes is selected as -60dB from 0dB reference value. From these potential modes a set of frequencies, damping and initial amplitudes of sinusoids are stored as reference features. As Figure 3 indicates, only those tracks are plotted that are above -60dB threshold.

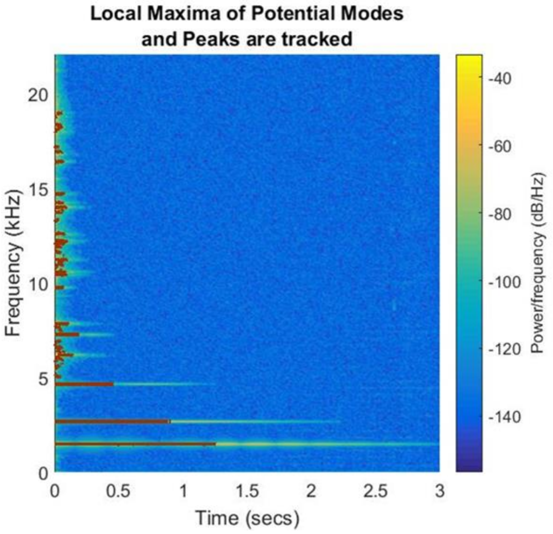

Figure 3: Tracked potential modes from audio clip

3.2 Parameter estimation and sound synthesis

Parameter estimation algorithm is designed to extract the best material parameter of a virtual object during its interaction. To compute the accuracy of the parameter estimation algorithm, we first created different graphical virtual objects (i.e., glass, metal and wood materials with known material parameters) and then excited with a unit impulse signal. The excitation modes are detected during the run-time interaction of virtual objects in order to synthesize sound. The synthesized sound is subjected to parameter estimation process. The estimated parameters and the parameters known beforehand for a wine glass are compared and the results of their relative errors are shown in Table 1.

Relative error in real and estimated material parameters	
α	8.64e-5
β	3.02e-5
γ	1.29e-6
σ	2.45e-5

Table 1: Relative errors in estimated parameter and parameters known beforehand for wine glass.

Based on these material parameters [8], the salient features are extracted for virtual object in the form of frequencies, damping and initial amplitudes as described in Equation 11.

$$\left(\lambda_j^o,\ a_j^o\right) \xrightarrow{\{\alpha,\ \beta,\ \gamma,\ \sigma\}} \left(\underline{\boldsymbol{f}}_i, \underline{\boldsymbol{d}}_i, \underline{\boldsymbol{a}}_i\right) \tag{11}$$

Figure 4 shows a comparison of reference features and extracted features in form of power spectral density estimation for a wine glass.

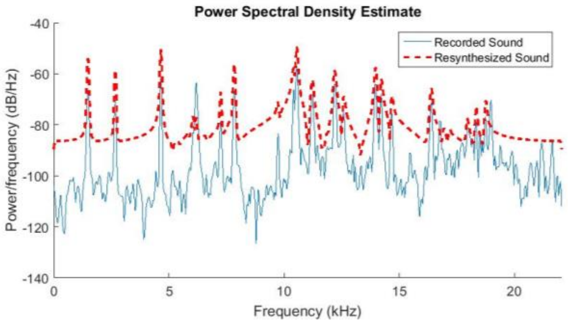

Figure 4: Comparison of modes for recorded sound and synthesized sound from frequency spectral analysis

Finally, using extracted features ‘$\underline{\boldsymbol{\varphi}}_i = (\underline{\boldsymbol{f}}_i, \underline{\boldsymbol{d}}_i, \underline{\boldsymbol{a}}_i)$’ in Equation 10, the virtual impact sound for a wine glass is synthesized. A residual compensation is performed to incorporate the non-harmonic part in final synthesized sound. Figure 5(a-c) shows spectrograms of original recorded example sound clip and synthesized sound.

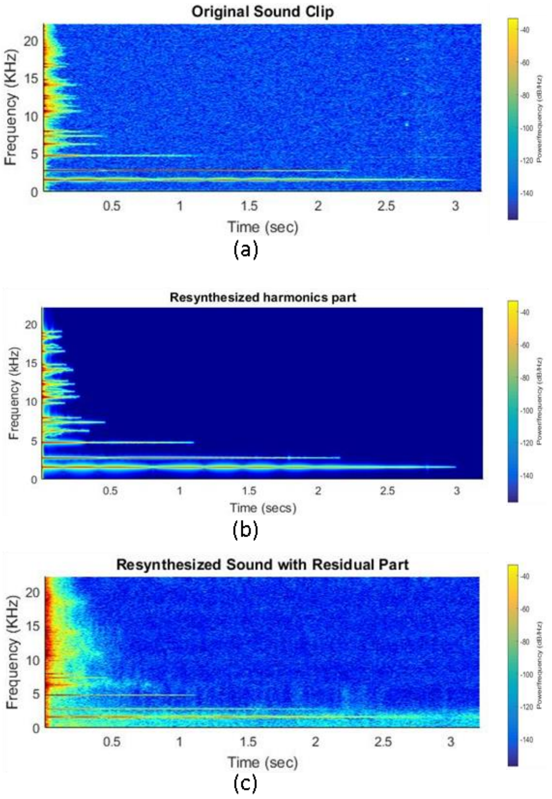

Figure 5: Spectrogram of a) recorded sound, b) synthesized sound and c) residual compensation of synthesized sound

4 DISCUSSION AND CONCLUSIONS

This study described and presented a physically based sound synthesis technique based on recorded sound of real objects. Using feature extraction and linear modal analysis, we estimated material parameters that represents the virtual object’s characteristics in synthesizing sound for graphical virtual environments. A residual compensation method is used to minimize the differences between real object sound and synthesized sound and incorporating the non-harmonic and noisy part in final synthesizes sound. In this study we showed an effective application of physically based LMS technique, however, there are certain limitations that are required to be considered. In LMS the homogeneous mass distribution is considered, whereas, in real scenarios the objects of different geometries does not necessarily possess homogeneous distribution of mass or density. Secondly, this approach work well for solid rigid bodies and may fail to recreate realistic sound for thin shells and deformable objects. Further studies are required to consider the non-modal approaches in order to achieve more realistic virtual sound for interactive deformable and thin shell bodies in run time applications.

ACKNOWLEDGEMENTS

This work was supported by the Global Frontier R&D Program on <Human-centered Interaction for Coexistence> funded by the National Research Foundation of Korea grant funded by the Korean Government (MSIP) (2013M3A6A3079356).

REFERENCES

[1] X. Serra, "Musical sound modeling with sinusoids plus noise," *Musical signal processing*.

[2] G. Durr, L. Peixoto, M. Souza, R. Tanoue, and J. D. Reiss, "Implementation and Evaluation of Dynamic Level of Audio Detail," in *Audio Engineering Society Conference: 56th International Conference: Audio for Games*, 2015.

[3] C. Zheng and D. L. James, "Harmonic fluids," *ACM Transactions on Graphics (TOG),* vol. 28, p. 37, 2009.

[4] C. Picard, N. Tsingos, and F. Faure, "Retargetting example sounds to interactive physics-driven animations," in *Audio Engineering Society Conference: 35th International Conference: Audio for Games*, 2009.

[5] J. F. O'Brien, "Synthesizing sounds from physically based motion," in *ACM SIGGRAPH 2001*, 2001, p. 59.

[6] M. Imran and J. Y. Jeon, "Determination of material parameters to recreate realistic audio quality for sounding materials in virtual sound reproduction based on modal analysis," *The Journal of the Acoustical Society of America,* vol. 137, pp. 2330-2330, 2015.

[7] D. B. Lloyd, N. Raghuvanshi, and N. K. Govindaraju, "Sound synthesis for impact sounds in video games," in *Symposium on Interactive 3D Graphics and Games*, 2011, p. 7.

[8] Z. Ren, H. Yeh, and M. C. Lin, "Example Guided Physically Based Modal Sound Synthesis," in *ACM Trans. Graph*, p. 16.