Virtual sound generation by linear modal synthesis based on recorded audio examples

Imran Muhammad, Jin Yong Jeon

Department of Architectural Engineering, Hanyang University, Seoul, Korea

Summary

Linear modal synthesis (LMS) has been investigated to synthesize virtual sound of rigid bodies. The existing methods to synthesize realistic sound of virtual objects lack the real-time computations of perceptually satisfactory material parameters. In this study, a novel approach is used for the estimation of the material parameters that record the intrinsic quality of interacting materials and to extract significant features from the recorded impulse responses. A parameter estimation technique based on psychoacoustic principles is developed that incorporates material features to search the best material parameters for LMS. The synthesized audio produced from LMS shows the same perception of material as a recorded sound.

PACS no. 43.72.Ar, 43.40.At,

1. Introduction[1]

In recent year, investigations in sound synthesis models for virtual environments have played a significant role especially for automatically produced sounds for various types of object interactions and situation. However, incorporating good quality audio to a visual cues and animated simulations, it remains a challenging task to portray the real-world complements. The basic challenge lies in emulating real world sounds accurately with the synthesized sounds because of inappropriate information of real sound effects. As a consequence, the audio-video perceptual cues lose its plausibility. Moreover, there are certain discrepancies in synchronization of audio-video during rendering process that create disparity among different audio and visual cues and hence produces a poor virtual perception. Traditionally, Foley artists manually record and edit the sound effects to a visual scene prior to the synchronization with video games, animated scenarios and movies with quite adequately satisfactory results but these procedures cannot be applicable to interactive environment.

In this paper an efficient sound synthesis technique is applied that produces the real and high quality of audio by manipulating the physically based modal synthesis process. In this approach, a single audio clip is required to be recorded as an example sound and from this recorded sound we estimate the intrinsic salient parameters of material that are later on employed as feedback in modal analysis. Once these parameters, such as stiffness of material, damping coefficients, and mass density are determined various kind of modes are produced differently for different interactions of bodies for different geometries in real time.

2. Methodology

The basic structure of the methodology is explained in Figure 1. As a first step we record one audio impulse response of one kind of material and then estimate material parameters that are used for modal synthesis in order to generate sounds. These synthesized sounds reflect the same sense of perception as original recorded sound.

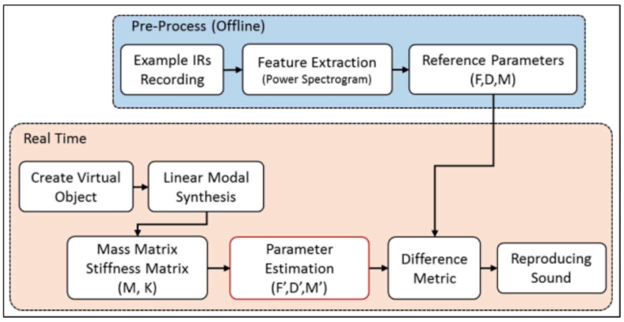

Figure 1. Flow of Methodology

Once the impulse response is measured for a certain material, it is subjected to feature extraction process like damped sinusoids of different frequencies, damping coefficients and initial amplitudes. The data set of these features is utilized in estimation process of material parameters for a virtual objects. For every impact of the virtual objects, the estimated parameters are used for the generation of virtual sound. The difference metrics in terms of psychoacoustic are calculated by a comparison of perceptual similarity between measured sound and virtually generated audio of objects. As an implication these parameters can be transferred to different object of varying geometries and shapes.

3. Modal Synthesis

For real time application much work has been done on modal synthesis for sound rendering applications. The modal synthesis processes are independent of any recorded sounds to create virtual sounds generated by interactions among virtual objects. In this way modal analysis does not need manual synchronization of the audio-visual events and the synthesized sound reflects a compromising variations of interactions among object and also their geometries [1]. However, for complicated scenarios, where various material interactions are involved, the computational cost and efficiency of parameter selection is affected by modal analysis.

In modal synthesis objects having same material properties produce same material sound such as sliding metallic box, rolling metallic ball and striking metallic plate sounds alike in similar fashion [1]. In contrast if we determine the material properties from recorded audio these properties can be transferred to different objects of varying sizes and various shapes.

Resonant mode are generate when objects collide with each other, called resonant modes with decaying amplitudes of mode vibrations in time. Lower frequencies modes have lower decay rate as compared to the high frequency modes. In modal synthesis we simulate the vibration of the resonant modes to reproduce the sound. Mathematically, the synthesized signal x(t) can be computed as [2], and shown in Equation 1.

$$x(t) = \sum_{m}^{M} g_m A_m(t) \, sin\big(2\pi f(t) + \varphi_{0,m}\big) \qquad (1)$$

Where M is the number of modes and g_m, $A_m(t)$, f_m, and $\varphi_{0,m}$ are the gain, amplitude envelope, frequency (Hz), and initial phase of each mode m, respectively.

3.1. Computation and Analysis

Computations of modes is carried out by short Fourier transformation (STFT). The audio recorded example is split into time windows and a spectrogram was computed. The modes are indicated as strong peaks in the spectrogram that persist over time at the same frequency [2]. The criteria of selection of the modes is to identify the strongest peaks for specified time window over a given threshold as defined by user in the spectrogram. The time slice is normally selected near the onset of the audio where high frequency modes have not fully decayed. The recorded audio of metallic strike and its spectrogram is shown in Figure 2(a). The detected and modeled modes are shown in Figure 2(b).

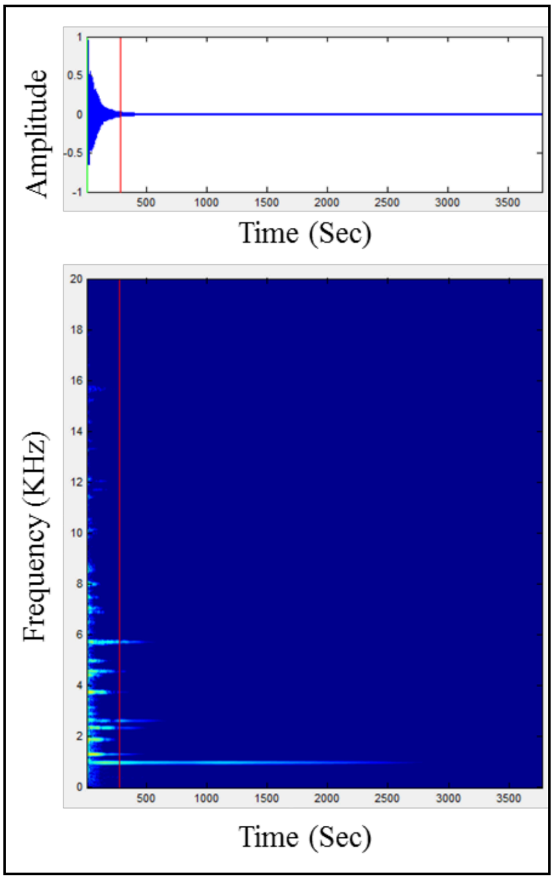

Figure 2(a). Recorded Audio for Metallic to Metallic strike and the spectrogram.

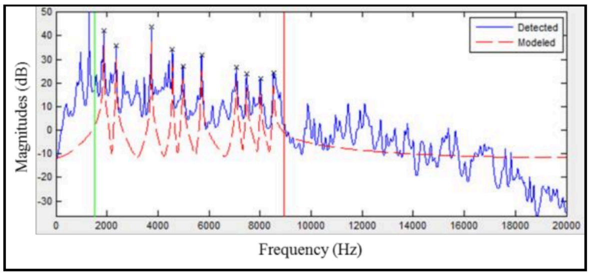

Figure 2(b). Detected modes form the Audio sample and modeled modes for virtual sounds. Modeled modes can be varied for recreating different sound effects.

After identification of the modes, the features are detected in terms of frequency, damping and amplitude envelope for that particular time slice. These features are termed in this paper as reference dataset.
The estimated sound for virtual object of the same size and shape as that of the original object is resynthesize using the reference data set. Tetrahedral mesh is formed to compute the mass matrix $\boldsymbol{M}$ and stiffness matrix $\boldsymbol{K}$ assuming the material is homogeneous by using finite element method with initial values for Young's modulus, mass density, and Poisson's ratio, $\boldsymbol{E_0}$, $\boldsymbol{\rho_0}$, and $\boldsymbol{v_0}$. Form these values we calculate Eigen value that further are used for computing the estimated values of the estimated parameters for virtual object excitations to generate the sound.

4. Conclusion

In this study we have presented a system for virtual sound synthesis of impact in run-time application such as video game and virtual environments interaction. Modal synthesis analysis technique is employed based on recorded audio IR. Feature extraction form recorded audio has been utilized in parameter estimation of the virtual object sounds and in the Linear Modal Synthesis. The estimated parameters can be applied to various interaction of the objects in virtual environment. Further work is required for the psychological evaluation for finding the perceptual similarities between the recorded sound and synthesized sound. Furthermore, work is required for handling the sliding and frictional sound that is of great interest.

Acknowledgement

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] Ren, Zhimin, Hengchin Yeh, and Ming C. Lin. "Example-guided physically based modal sound synthesis." ACM Transactions on Graphics (TOG) 32.1 (2013): 1.
[2] Lloyd, D. Brandon, Nikunj Raghuvanshi, and Naga K. Govindaraju. "Sound synthesis for impact sounds in video games." In Symposium on Interactive 3D Graphics and Games, pp. PAGE-7. ACM, 2011.