Binaural rendering using HRTF

Niroshan Vijayarasa - Audio Signal Processing Course - 23.12.2018

Abstract— **When we are listening to music using headphones, the source is panned in 180 degrees. Using Head Related Transfer Functions(HRTF) this mini-project construct a program to listen to music in 360 degrees. The input of the program will music, and the user will be able to place the source in function of the distance and direction of the listener. The program will then simulate the overall music by into taking in account where they come from. The result of the project is a web-application written in JavaScript which is hosted on internet.**

I. INTRODUCTION

The human uses the ears and the head movements to localize the source of the sound and the interpretation of where it is coming from is mostly done by the brain. To perceive the spatial image, our auditory system uses the following cues: interaural time difference and level difference. It addition, the movement of the head is uses to resolve the cone of confusion.

In this project, we will use those features to simulate the localization of the auditory objects. From an input source placed virtually in a free-field space, we simulate the location of the object through the headphones. This simulation has a dynamic property which is that the auditory object can be moves in real-time and the rendered audio signal will change according to the displacement of the source.

As you can see in the figure 1, the blue circle represent the listener and the black square is the auditory objects. The simulation that into account the angle of arrival and the distance from the listener.

The challenge that is posed by this problem is multiples: First, Localization of the sound is done by the ears and the head movements. The latter one is used to resolve the cone of confusion. Thus, using the head-phones, we won't be unable to use the head movements. Second, with the various shape of the ears that the human has, the impulse response that represent them are numerous and this leads to a resolution of the problem that is experimental. It means that the solution might not work for some and for other it can work perfectly.

II. REVIEW OF EXISTING TECHNIQUES

The resolution of the problem can be divided in two steps: the first part is to resolve the angle of arrival of the source, and the second part concentrate on how to add the distance component in our simulation.

As the title suggests we will use the head-related transfer functions(HRTF) to resolve the angle of arrival. For the explanation, our ear is composed of the internal and external part. When we are listening, the final destination of the sound is the basilar membrane. Before hitting this membrane, the sound reflect on numerous part of our ear such as the pinna, tympan and those reflections are used to compute the delay and the intensity level which resolve in the localization of the sound.

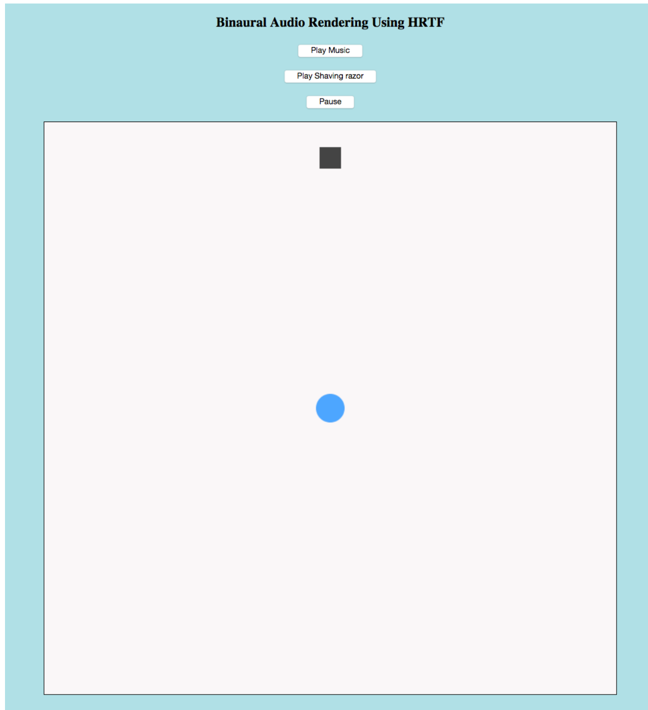
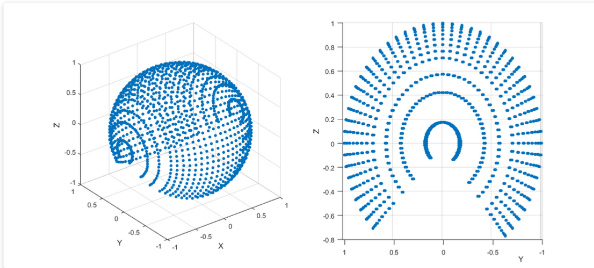


Fig. 1. The user interface of the web application

In this exercise, we model this phenomena using impulse responses which we call, the Head-related impulse response(HRIR) and the HRTF are the Fourier Transform of the HRIR. Deriving the HRIR directly is a difficult task because the impulse response varies from a person to person and that he computation of the impulse response is complex to do in a computer.

To counter this problem, we use the CIPIC Database. It is a database that has a collection of impulse responses computed in function of the elevation and azimuth angle of the source to the listener. More precisely, the collected impulse response are measurements which involved 50 subjects. Each one of them is placed in the center of an anechoic chamber with microphones place in the ear. A source is place in 25 different azimuths and 50 elevations angle in order to measure the HRIR. We therefore choose this database because it is a popular and work well for the project.

The database has finite number of HRIR and we want a localization that is panned in 360 degrees 2. This problem is address using interpolation technique. The idea is to create



Visualization of points at which HRIR from CIPIC are recorded.

Fig. 2. CIPIC database points

a general impulse response using the measured one.

The interpolation technique is linear and the explanation is the following: for an angle θ which is defined from the front of the head to the source position, we find the two closest angles θ_{-} and θ_{+} which englobe θ . We compute the normalized distance from the two angles, which we define as the weight of out interpolation, we call them α and β. The impulse response for the /theta angle is given by $I_{\theta} = \alpha * I_{\theta -} + \beta * I_{\theta +}$ where $\alpha + \beta = 1$

The second part consist of adding the distance property in our simulation. A sound is modeled as a point source and the amplitude of the sound decease inversely proportional to the distance. We use this model to give the feeling of the distance in the simulation. Define: $s(t)$ as the source and $r(t)$ as the listener. We therefore relate them as $r(t) = s(t)/r$ where r is the distance between them.

III. Algorithm Description

The output of the project is a web-application written in HTML, CSS and Javascript and the implementation of it was done in three step. The first step is the pre-eliminary work which consist on handling the CIPIC database using Matlab script. The second step is to create the web-application back-end where the pipeline of it is created with the Web Audio Api. The last step is the development of the GUI that represent the front-end of the application.

The impulse response found in the CIPIC database is in Matlab format. Testing the database is an important step for the project, it helps to know if the database is fitted for the problem and to understand what are the needs for the implementation.

The written Matlab script is able to run a simulation given an azimuth and elevation angle and it was used to choose which subjects impulse response from the 50 subjects we will choose for the web-application. Thus, We choose the subject number 3.

The core of the project is the back-end of the web application. We choose to implement it using Web Audio API. It is a javascript library that is used to deal with audio signal. The building of the web-application is done by creating and putting together audio components in order to create a pipeline.

The figure **??** represent the audio components that was used to create the simulation. The SourceNode is the input sound that will be localized by the listener. We use a mono signal that is centered and preprocessed with Audacity. Then, this sound is fed to the panner which is composed using two convolver.

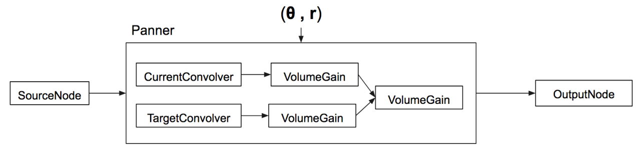


Fig. 3. Program pipeline

The panner is the audio element which produce the simulation of the localization. The reason why we use the two convolver is the following: the current convolver is the one which is actif during the listening and the target convolver is the one that is used during the displacement of the auditory objects. Transition between two position of the source needs to be smooth and allowing two convolver will do the job. The transition is done as the following: when the source is moving we convolve the sound using the target convolver with the new position. Then, we decrease the volume of the current convolved in order to increase the volume of the target convolver. This volume fade last 25 ms. Finally, the volume gain that is connected to the output is used to simulatie the distance property. The final volume gain in the panner is computed using the distance ratio r and directly connected to the output.

Finally, the front-end is a GUI implemented using simple HTML & CSS and Javascript. Three button are present, the first consist on playing the music, the second is to play the sound of a shaving razor and the last button enable to pause the sound. To interact with the canvas, we can drag and drop the source sound (represented as the black box) where ever the user wants around the listener. The angle and the distance is then computed in order to give as argument to the back-end.

IV. Results and Discussions

Note that there is two sound source that we can alternatively choose, the music and the sound of the shaving razor. The latter signal was chosen in order to improve the immersion of the simulation and to propose diversity to the user interaction. The sound that is generated gives a good simulation but can be improved.

The interpolation is done in the time-domain using weighted combination of existing impulse responses therefore we dont allow phase change in the resulting impulse response. It means two impulse responses can have difference delay cues, doing the interpolation as presented will not add or subtract existing delays. The way to resolve this problem is to do the interpolation in the frequency domain.

V. Conclusions

As future work, the above presented interpolation technique can be implemented and compare with the actual one.

In addition, we can augment the simulation by allowing angles that cover all the sphere.

REFERENCES

[1] V. R. Algazi, R. O. Duda, D. M. Thompson and C. Avendano, The CIPIC HRTF Database, Proc. 2001 IEEE Workshop on Applications of Signal Processing to Audio and Electroacoustics, pp. 99-102, Mohonk Mountain House, New Paltz, NY, Oct. 21-24, 2001.

[2] Audio Signal Processing and Virtual Acoustics, Christof Faller and Mihailo Kolundija, Chapter 7

[3] Sound Localization - Wikipedia

[4] Implementing Binaural (hrtf) Panner Node with Web Audio Api -

[5] http://jsfiddle.net/KZ99q/