DESC 9115

DIGITAL AUDIO SYSTEM

ASSIGNMENT 2

Name:  Syed Azhar Alatas Syed Ahmad
SID:    310078237
1.0 Introduction

Digital signal processing (DSP) is distinguished from other areas in computer science by the unique type of data it uses: signals. In most cases, these signals originate as sensory data from the real world: seismic vibrations, visual images, sound waves, etc. DSP is the mathematics, the algorithms, and the techniques used to manipulate these signals after they have been converted into a digital form. This includes a wide variety of goals, such as: enhancement of visual images, recognition and generation of speech, compression of data for storage and transmission, etc. The DSP technique also widely use in sound processing technique such delays, spectral processing and spatial effects. A 3D positional audio is one of the techniques in DSP.

1.0 Problem Description

More and more 3D movies are making their way to the big screen these days. Could 3D audio be the next technology production companies roll out in order to lure customers back to cinemas and away from free downloads over the internet? Millions of dollars in tickets sale of 3D movies indicates that ‘3D is the story’. A very significant number of theatres are being equipped for future 3D releases suggesting that the movie format could be here to stay this time. In terms of sound system, however the 3D sound technology in cinema did not fully perform in order to match the 3D visual image with the spatial effects of the sound even though the spatial sound technology is well develop to enhance the performance of the visual images. There are two main difficulties of the present cinema regarding the 3D sound:

1. Audience sitting area
   Audience who sit in the middle of the sitting area in cinema will experience the sound from the movie better than the audience who sit at the rear, front or at the corner. It seem that the sound system cannot fully work for every audience in the cinema

2. Quality of sound system
   Every cinema will have different quality of sound system and almost impossible to experience the same sound quality.

The audience’s head during watching the movie at the screen in the cinema is always not static. Anything can happen to the sounds that come from the speaker to the eardrum. These sounds are subject to noise, the cinema’s architectural acoustics, delays, etc or in other words, the sound are widely modified by the environment. A part from that, given that speakers can only create a convincing effect for a small number of seats in the cinema at the so-called ‘sweet spot’ such as the middle seat. The precedence affects also one of the present problems in the cinema regarding the sound effects since position of every listener is different. This is because our hearing system is really sensitive to the direction of the first incoming wave-front, when the identical signals are
radiated from a set of stereo loudspeakers, stepping to one of the centre causes the transformation of directivity.

3.0 Specification

Given the success of 3D film and the willingness of audiences to wear 3D glasses, it can be imaginable that the audience would be happy to wear headphones at the movies for a 3D audio effect. The problem of the 3D sound using loudspeaker in cinema can be overcome by using headphones at the cinema. This would allow for the ultimate surround sound experience, given that speakers can only create a convincing effect for a small number of seats in the cinema at the so-called ‘sweet spot‘. Movies could also be played in many different languages at the same time. One of the initial theories for explaining sound source localization has been labeled the *duplex theory*. This duplex theory attributes our localization abilities to two different interaural cues: interaural time differences (ITD) and interaural intensity differences (IID). An ITD is defined as the difference in arrival times of a sound's wave-front at the left and right ears. Similarly, an IID is defined as the amplitude difference generated between the right and left ears by a sound in the free field. In general, a sound is perceived to be closer to the ear at which the first wave-front arrives, where a larger ITD translates to a larger lateral displacement. An aliasing problem can occurs above 1500 Hz, and the difference in phase no longer corresponds to a unique spatial location.

The common belief is that the shape of the head, torso, and especially pinnae affect the spectral contents of a sound as it propagates from free space to the inner ear. Additionally, because of the irregular shape of the upper body and pinnae it is theorized that these filtering effects would be spatially dependent, varying with the location of the sound source. This filtering effect is commonly referred to as a Head-Related Transfer Function (HRTF). Because the HRTF would vary with elevation and azimuth, this would create a unique HRTF for each sound source location. Each individual has unique physical shapes and characteristics, so each individual will also have a unique set of HRTFs. The computed HRTFs for a location would contain both the ITD and IID cues and the spectral filtering effects, primarily due to the pinnae. Because the elevation of the sound source relative to both ears will be the same, the spectral filtering effects would also be expected to be very similar. With this spectral cue being roughly the same at both ears, we don't have a significant interaural frequency difference to compare between the ears. This explains the empirical observation that, lacking visual cues, we are much better at localizing sound in the azimuthal plane than in the elevation plane. The important information such as spatially related spectral cues or the source location can be captured via an impulse response. This impulse response is termed the Head-Related Impulse Response (HRIR). Convolution of an arbitrary source sound with the HRIR converts the sound to that which would have been heard by the listener if it had been played at the source location, with the listener's ear at the receiver location. This is the concept of fundamental idea of spatialization. If the audience receives the sound material via a stereo headphone, we can reproduce most of the cues that are due to the
filtering effect of the pinnae-head-torso system, and inject the signal artificially affected by this filtering process directly to the ears that will give the impression of the 3D sound.

4.0 Implementation

Right now all over the world, there is no single cinema or theatre that uses HRTF format and headphone to play the 3D sound for the movie. If we are the first that introduce this kind of concept, the product, the cinema and the 3D movie with 3D sound could be a hit and success. As discussed earlier, given the success of 3D film and the willingness of audiences to wear 3D glasses, it can be imaginable that the audience would be happy to wear headphones at the movies for a 3D audio effect. It is possible to deliver multi-channel surround sound via stereo headphone reproduction. There was already a practice on this but not in the cinema world. Wireless headphone, transmitter and decoder (in one housing) will be installing at each seat for every audience. The integrated Dolby surround, Dolby Pro Logic and Logic 7 decoders will instantly grant the audience with access to the multi-channel audio with sophisticated detail and a truly impressive imaging and spaciousness. The headphone’s transmitter; with a long range will allow the audience enhanced freedom to choose the sitting area and can be use in the large based cinema. The headphone will also have batteries that can offer acoustic freedom. These headphones can actually work better than speakers. The reason for this is that we only got two ears. The way we tell whether a sound's in front, behind or above us, rather than just to our left or your right, is by processing the complex differences in phase, time delay and frequency balance that a re imparted to differently located sounds by nearby objects (like walls), and by the sonic characteristics of the head.

Convolution process that is a mathematical way of combining two signals to form a third signal will be use. In linear systems, convolution is used to describe the relationship between three signals of interest: the input signal, the impulse response, and the output signal. An input signal, \(x[n]\), enters a linear system with an impulse response, \(h[n]\), resulting in an output signal, \(y[n]\). In equation form: \(x[n] \ast h[n] = y[n]\). Figure 1 depicts how the concept of the convolution will be used. Implementation of this application is not complicated. As the discussed earlier that the process of localizing a sound source convolving with the HRIR, each sound source such as voice, sound effect can be convolve with HRIRs that have certain directivity information such as azimuth and elevation. This process can be done earlier during the post production of the movie. However, the generic HRIR will be use rather than individualized. From the generic HRIR, average HRIR will be use. When we got HRTF-massaged two-channel audio already, for instance when we are watching the movies, headphones are obviously the best way to get the sound into the head. There's no way for the speakers to do the job as well, because there's no way for them to stop each ear hearing the sound that's intended for the other. The HRTF algorithms will fake the effects of the pinnae, the head and various listening environments, and by injecting the sound straight into the ear canal can produce the impression of spatial audio sources. Figure 2 depicts the process.
5.0 Evaluation

Since today most cinemas choose loudspeakers rather than headphones, it is worth mentioning how the growth of 3D sound content in cinema may affect the headphone market. Technologies especially electronics will move faster from one stage to another stage. Up to now, in a market using conventional stereo at best, audience might think using headphone while watching the movie can improve the privacy. One issue of using headphone is that headphone commonly has excellent bass response, but their teeny little drivers don't move bulk air like a big speaker. This can be overcome by installing high quality drivers in the headphone. There is another issue of applying the headphone and HRTF concept applying to cinema world. During watching the movie, audience head are not always static to the screen. This can be overcome by using active head tracking system. By using the active head tracking system, the audience will have the same surround sound experience, or the sound stays ‘glued’ to the screen although their head is not to the screen. A part from that, the sound that comes from the headphone can be subject to noise.
Noise can be generating from the cabling, during the transmission or from the equipment itself. This can be overcome by using the noise cancellation headphone. The headphone must be a high quality type and system with the features as discussed earlier. Externalization is another issue for 3D spatial sound. One useful technique which can enhance the localization of the sound sources when using headphone is decorrelation. This is because the signals arriving at the ears are completely decorrelated in natural situations. Headphones with Mobile Surround or Dolby headphone technology will work perfectly for localization. It will require much more investment in the headphone system. But it is worth for investment as the 3D audio sound technology is gathering pace. Possible collaboration with Hollywood movie maker, this first technology in the world for cinema sound system using headphone with HRTF concept, will give a huge success and profit. It will give new dimension of surround sound system using headphone to the technology world and to the users.

6.0 Extra Credit

Constructing an accurate model of human sound localization is a complex task. When the listener estimates the location of an incoming sound many factors other than the stimulus itself can affect the judgment. Listeners are likely to incorporate visual information into their interpretation of the sound, which can alter the listener’s perception of the auditory spatial information. Moreover, familiarity with the environment and prior knowledge of the type of sound and its amplitude and frequency characteristics can help to improve the localization performance of the listener. A positionally stable 3D sound image comes mostly from a wavelength-based interaction of the arriving sound with the head and the visible part of each ear. This is especially true for higher frequencies, say from 1500 Hz up. A stable image may not be possible if the listener's head moves by more than part of a wavelength. At 12 KHz, for example, a wavelength is about an inch. There's a common-sense way to see why head position change can have an effect. It's reasonable that a head position change larger than their detail size (which would blur or destroy their acoustic effect) could also blur localization in a loudspeaker sound-field. The performance of the system can be assessed by the audience that although their head move, by using headphone the sound deliver will be a stable binaural signal. The sound should be heard outside the head or called having “out-of-headedness”. In other words the sound seems to come from outside where it should, not from inside the headphones and especially not from between the audience’s ears. This is because the left and right binaural or 3D sounds are precisely correlated with each other. The user should be hear a different experience of sound in free field listening, normal headphone listening and headphone listening with binaural or 3D sound that are using HRTF concepts. For the user to hear the difference, in free-field listening, the sound signals arriving at either ear can differ from one another greatly for sound waves produced by the natural acoustic sources. In normal headphone listening with typical headphone playback presents sound signals to the ears that do not include the acoustic effects of the HRTF. Listening through headphones with binaural processing (using HRTFs) presents transformed sound signals to the ears that restore these natural acoustic effects. It is well known that the HRTF are unique for each person. The major issue is that each person’s direction-dependent set
of HRTF is distinctive, and substituting one person's for another's tends to produce vague and inaccurate localization. This is mainly because the physical form of the external ear (especially the pinnae) varies substantially between individuals (both in terms of shape and size).