RESEARCH ARTICLE OPEN ACCESS

A Sound Project Using Python

Dhananjay Sharma, Radheshyam Madge, Rishabh Mishra and Nissim Shewade
Department of Computer Engineering Sinhgad College of Engineering University of Pune, Pune India

Abstract
This paper provides the implementation of various digital audio effects (DAFXs) as a combination of user defined parameters and input sound signal.An approach to implement various effects like delay based effects, spatial effects, time varying effects and modulators is provided.A unique listening environment is provided using 3-D spatialization and localization, simulated surround sound, dialogue normalisation, dynamic range control and down-mixing.An attempt has also been made for music and voice separation to provide karaoke effect to the sound.
All the implementations are provided in python which is widely adopted, open source and general purpose programming language and has a vast array ofcode libraries and development tools, and integrates well with many other programming languages, frameworks and musical applications.
Keywords- Digital audio effects; Dialog normalization; Dynamic range control; Audio downmixing

I. INTRODUCTION

THIS paper has been written in support of a sound project which largely encircles the domain of digital audio effects – DAFX as an acronym. The Python programming language has been chosen for development of this project due to its apprehensive provision of sound API's, tools, libraries and support for audio signal processing and its comprehensive ability to integrate with other programming constructs. At an abstract level this project comprises of two modules - audio effects and dialog normalization coupled with dynamic range control.
The concept underlying dialog normalization and dynamic range control goes hand-in-hand. Consider for example a case of TV channels where each channel sounds different although all of them have been set at the same volume. Likewise, a portion of an audio signal may also sound different from another one although both portions are played at the same volume. Hence, the challenge here lies in controlling and modifying values of chunks (data packets) in each portion to be not more than, so also not less than a threshold value, the end result being both portions sound similar when played at the same volume. A code for the same has been implemented in this project.

Digital audio effects (DAFX) are systems that modify audio signals. These transformations are made according to some control parameters the permits and delivers output sounds. The audio effects module being an extensively vast module in itself can be further classified into modules characterized by the inherent property of audio signal that is modified by each effect as:

- Delay based effects – chorus, flanger, vibrato etc.
- Spatial effects – reverberation, panning, 3D effect for headphones using hrtf
- Time varying filters – wah-wah
- Modulators – tremolo
- Karaoke using audio downmixing
- Other exciting effects

In order to reduce the restrictions on the type of audio file that can be scanned as input, a provision for audio transcoding (i.e. mp3 to .wav conversion) and compression of a .wav file has also been provided for. A dedicated attempt has been made to create and integrate a comprehensive listening and sound manipulating environment in a single project as standalone software which provides for uniqueness to the project.

II. DIALOG NORMALIZATION AND DYNAMIC RANGE CONTROL

In accordance with the example of TV channels that has been mentioned previously in this paper, in the introduction section we hereby throw some light on this concept and utilization in sound based application. The dialog normalization coupled with dynamic range control algorithm, basically checks the value of amplitude of each chunk (data packet) in an audio signal and reduces it to the threshold value if it is greater than the threshold, and increases the value to the threshold value if value of the chunk is less than the threshold. Hence, at the same volume, all portions of the same audio file which sounded differently now sound similar. This

concept has been implemented using the following equation:

If $x(n)>x_h(n)$,
$y(n)=x(n)- \delta(n)$
And
If $x(n)>x_l(n)$,
$y(n)=x(n)+ \delta(n)$.

The application of this concept can be found in the scenario where, a TV channel sound sounds normal, but advertisement on the same channel sound louder at the same volume. Using this concept, the advertisements can be made to sound only as loud as the TV channel sounds normally.

III. AUDIO EFFECTS

Audio effects is an exciting area in the field of audio signal processing which throws light on the technique used to manipulate an audio signal by modifying one or more of its inherent properties like time distribution, spatial distribution, frequency etc. A well-defined comprehensive algorithm has been used to create each effect only after thorough reference of many papers published at various institutions in this regard. Specification of each effect and the corresponding algorithm used has been provided in this paper.

Delay-based effects:

These category of audio effects are created by introducing a time delay in between two samples of the audio signal. Chorus, flanger and vibrato are the delay-based effects which have been implemented in this project.A chorus effect simulates the effect of multiple instances of the signal which are close in frequency playing along with the lead instrument in unison. It is achieved by incorporating a delay in the given sample using the transition function $\delta(n)=x(n-k)$.The delayed samples of signal are then transmitted together with the original signal, i.e. original signal imposed with delayed samples of original signal.Hence the final equation for a chorus effect is concluded to be $\delta(n)=x(n)+x(n-k)$.

A flanger effect is obtained when two identical audio signals are mixed with one signal delayed by a small and gradually changing period. The time delay here is smaller than 20ms. Peaks and notches are produced in resultant frequency spectrum. Feedback is used to enhance the intensity of peaks and troughs. The algorithm and equation used to implement a flanger is same as that of chorus effect. The only difference being a varying delay of much less magnitude is induced to get a flanger.

Vibrato is a regular fluctuation in pitch, timbre and/or loudness of the audio signal. An acoustic signal here, can be visualized as a set of partials, composing the voiced sound, whose amplitude and frequency vary along time in a particular way. In case of vocal vibrato, the position of these partials, i.e. their frequency variation is harmonically related to the fundamental frequency variation and is expected that all of them exhibit the same regular fluctuation. Vibrato effect is implemented using the equation $\delta(n)=x(n)+g.x(n-m)$. Here $\delta(n)$ represents transition function. Here g is the gain which retains a constant value between 0.3-0.7. Also m=t/fs is the sample frequency, y(n) output signal and x(n) is the input signal.

Spatial Effects:

Reverberation is the persistence of sound in a particular space after the original sound is produced. A reverberation, or reverb, is created when a sound is produced in an enclosed space causing a large number of echoes to build up and then slowly decay as the sound is absorbed by the walls and air. This is most noticeable when the sound source stops but the reflections continue, decreasing in amplitude, until they can no longer be heard. In comparison to a distinct echo that is 50 to 100 ms after the initial sound, reverberation is many thousands of echoes that arrive in very quick succession (.01 – 1 ms between echoes).Reverberation effect can be implemented with the help of impulse response.The impulse response is convulsed with input signal to get the reverberated output. Discreet convolution formula:

$$y(n)=\sum x(k)*h(n-k)=x(n)*h(n) \quad (k=-\infty \text{ to } k=+\infty)$$

Panning is the spread of a sound signal (either monaural or stereophonic pairs) into a new stereo or multi-channel sound field.A typical physical recording console pan control is a knob with a pointer which can be placed from the 8 o'clock dial position fully left to the 4 o'clock position fully right. This software replaces the knob with an on-screen "virtual knob" or slider for each audio source track which functions identically to its counterpart on a physical mix console.The pan control in audio gets its name from panning action in moving image technology. The audio pan control can be used in a mix to create the impression that a source is moving from one side of the soundstage to the other, although ideally there would be timing [including phase and Doppler effects], filtering and reverberation differences present for a more complete picture of apparent movement within a defined space. Panning effect has been implemented using the following equation:

We seek to obtain to signals one for each Left (L) and Right (R) channel, the gains of which, gL and gR, are applied to steer the sound across the stereo audio image. 'x' is mono input. This can be

achieved by simple 2D rotation, where the angle we sweep is:

A =[cos(angle), sin(angle); -sin(angle), cos(angle)]
[gL,gR]'=A.x

3D audio effect with headphones:

There are two models of implementing this effect. One is mathematical model and the other is HRIR convolution method of localizing mono sources in headphones that could successfully 'place' the source at the desired angle. Each of the methods had their advantages and disadvantages in implementation, but the end result seemed to be reasonably accurate for both. We have implemented the second one. The HRIR convolution method had the advantage of not making physical approximations, but rather measuring the exact impulse response for the appropriate angle. However, this method results in discrete locations for all the angles, and thus a continuous sweep would not be possible. In its current form, if an angle is not present in a recorded impulse response, the closest one must be used.

Time-varying Filters:

Wah-wah (or wa-wa) is an imitative word (or onomatopoeia) for the sound of altering the resonance of musical notes to extend expressiveness, sounding much like a human voice saying the syllable wah. The wah-wah effect is a spectral glide, a "modification of the vowel quality of a tone", also known as a band pass filter. The wah-wah effect is produced by periodically bringing in and out of play treble frequencies while a note is sustained. The word is derived from the sound of the effect itself—in other words, it is onomatopoeic.This effect has been implemented using the following equation:

$$\delta(n)=y_b(n)+y_h(n)$$

Here, x(n)=input signal, $F_1 = 2*\sin((pi*Fc(1))/Fs)$, $y_b(n)=F_1.y_h(n)+y_b(n-1)$, $y_h(n)=x(n)-y_1(n-1)-Q_1.y_b(n-1)$
Also, F1 = 2 sin(fc=fs), Q1= 2d and y(n)=output signal.

Modulators:

Tremolo is basically produces a trembling effect.Tremolo is a variation in amplitude which turns the volume of a signal up and down, creating a "shuddering" effect. This type of effect is often used by electronic instruments and takes the form of a multiplication of the sound by a waveform of lower frequency known as an LFO. The result is similar to the effect of rapid bowing on a violin or the rapid keying of a piano. In accordions and related instruments, tremolo by amplitude modulation is accomplished through intermodulation betweentwo or more reeds slightly out of tune with each other. This effect has been implemented using the following equation:

trem=(1+ alpha*sin(theta)) (alpha=constant=0.5)
y = trem*x

Karaoke using Audio Downmixing:

Audio downmixing algorithm essentially converts a stereo file (having 2 channels) to a mono file (having 1 channel). The structure and composition of a wave file attains special significance in the creation of this effect. All the vocal content is basically stored in the central part of a channel in a wave file. Hence, it has been observed that if two channels of a stereo file are attenuated separately and then one channel is subtracted from the other all the vocal content is filtered and all that remains in the wave file is instrumental music with only a slight stencil of vocals. Complete elimination of vocals is an aspect to be considered in the future scope of this project. Since we directly deal with channels here, there is no specific equation that describes implementation of this effect.

Other exciting effects:

In audio engineering, a fade is a gradual increase or decrease in the level of an audio signal. The term can also be used for film cinematography. In fade-out effect the amplitude of a sound decreases from start to end of sound file. A recorded song may be gradually reduced to silence at its end. Fading-out can serve as a recording solution for pieces of music that contain no obvious ending. In fade-in effect the amplitude of a sound increases from start to end of sound file. A recorded song may be gradually increase from silence at the beginning (fade-in).

The audio denoising effect is used particularly to remove the white noise present in a signal. During audio recording there is a possibility that the white noise present in the background may hamper the quality of recording. Hence in order to reduce the white noise denoising is applied. It doesn't reduce the noise directly to zero, but gradually eliminates it.

The fuzz audio effect comprises of complete non-linear behaviour, harder/harsher than distortion. It alters an audio signal until it is nearly a square wave and adds complex overtones by way of a frequency multiplier.

In overdrive effect, audio at a low input level is driven by higher input levels in a non-linear curve characteristic. Overdrive effects are the mildest of the three (distortion, fuzz & overdrive), producing "warm" overtones at quieter volumes and harsher distortion as gain is increased. A "distortion" effect produces approximately the same amount of

distortion at any volume, and its sound alterations are much more pronounced and intense.

Pitch shifting is a sound recording technique in which the original pitch of a sound is raised or lowered. A pitch shifter is a sound effects unit that raises or lowers the pitch of an audio signal by a preset interval. For example, a pitch shifter set to increase the pitch by a fourth will raise each note three diatonic intervals above the notes actually played.

Ring modulation is a signal-processing function in electronics, an implementation of amplitude modulation or frequency mixing, performed by multiplying two signals, where one is typically a sine-wave or another simple waveform. It is referred to as "ring" modulation because the analog circuit of diodes originally used to implement this technique took the shape of a ring. This circuit is similar to a bridge rectifier, except that instead of the diodes facing "left" or "right", they go "clockwise" or "anti-clockwise". A ring modulator is an effects unit working on this principle.

The robotic effect is widely used in movies (eg. star wars).The general speech signal is fed with some carrier wave to add robotic effect. The input is fed through band-pass filters to separate the tonal characteristics which then trigger noise generators. The sounds generated are mixed back with some of the original sound and this gives the effect.

In rotary speaker effect it appears as if the sound is rotating. The “Doppler” effect (varying the pitch based on the speed of moving towards and away from the sound source) will be familiar to you from ambulance sirens zooming past in the street! Rotary effects use the same phenomenon, but on a smaller scale, and combined with the tremolo (more later) effect.

The spectral panning effect consists in converting a mono sound signal into a stereo sound signal where each frequency component is placed to its own azimuth position between the loudspeakers, creating an spectral split effect.

In time stretching you can either stretch or compress the signal. The time period for which the audio signal lasts increases or decreases accordingly. It doesn't change pitch of audio.

IV. CONCLUSION

The audio effects implemented in this project have been observed to have considerable accuracy. The karaoke effect has a stencil of vocals left in the final output which ideally should be completely eliminated. Thus, complete elimination of vocals herein should be considered as future scope of this project. Also optimization in the implementation of dynamic range control should be considered in the future scope. All effects promised in this paper have been implemented and delivered, the accuracy of each varying to a certain extent.

REFERENCES

[1] UdoZolzer “**DAFX - Digital Audio Effects**” Copyright q 2002 John Wiley & Sons, Ltd ISBNs: 0-471-49078-4 (Hardback); 0-470-84604-6 (Electronic).

[2] IxoneArroabarren, Xavier Rodet, and Alfonso Carlosena “On the Measurement of the Instantaneous Frequency and Amplitude of Partials in Vocal Vibrato”IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 14, NO. 4, JULY 2006.

[3] UdoZ¨olzer “PITCH-BASED DIGITAL AUDIO EFFECTS” 5th International Symposium on Communications, Control and Signal Processing, ISCCSP 2012, Rome, Italy, 2-4 May 2012.