Justification and Theory for a Frequency-Specific Spectral Panning tool based on Duplex Theory

Niall Garry

University of Lincoln
15558790@students.lincoln.ac.uk
Supervisors: Prof. Chris Bingham, Dr. Matt Booth

ABSTRACT

The following paper is an investigation into duplex theory of sound localization and the justification for a spatial audio application based on the theory. Functions and filters are defined for the time/phase differences and the amplitude/spectral differences between each ear caused by sounds originating from a specific location relative to the listener. These transformations are applied within a standalone application created in MATLAB that can take a monoaural sound file, apply a specific spatial characteristic relative to azimuthal angle, elevation angle and head radius, and output a stereo sound file.

1. INTRODUCTION

As humans, we perceive sound in three-dimensions. Our hearing is a massive part of how we perceive the environment around us, with our brains often able to interpret sounds around us as having a location. This 'Spatial Audio' effect can be artificially recreated and played over headphones or loudspeakers.

For years, spatial audio has been a massive field of research due to its many applications; music production, virtual reality, cinema speakers etc. The phenomena can be recreated in a variety of ways; the most popular being by recording sound using specialised microphones or by artificially applying alterations to the sound within software, usually applying a known Head Related Transfer Function. The focus of this paper is to provide the theory and justification for a spectral 'panner' based on duplex theory that can apply a desired perceived location to a monoaural sound via completely computational means. An application is also created within MATLAB that uses this method to apply the effect to a sound, to provide the basis for numerous applications.

2. DUPLEX THEORY

Duplex theory, coined by Lord Rayleigh during his experiments with tuning forks in 1906 is a theory of sound localization. He concluded that there are two primary mechanisms at work for humans to locate sound; the detection of timing or phase differences between ears, and the detection of amplitude or spectral differences between the ears [1]. These timing and amplitude differences between the ears are named the Interaural Time Difference (ITD) and the Interaural Level Difference (ILD) respectively.

2.1. Interaural Time Difference

The Interaural Time Difference is caused by the distance between the two ears of the listener; therefore, it is calculated as a function of the head radius of the listener. Through simple trigonometry the following equation can be derived:

$$\Delta t = 2\left(\frac{a}{c}\right) sin\theta \tag{1}$$

Where a is the head radius and c is the speed of sound in air. Θ is the azimuthal angle, the horizontal angle from the centre front of the listener. However, previous research has found that the following formula fits experimental data much better [2]:

$$\Delta t = \left(\frac{a}{c}\right)(\theta + sin\theta) \tag{2}$$

2.2. Interaural Level Difference

The Interaural Level Difference is currently a strong point of discussion amongst researchers. This is because a significant amount of the level difference is caused by the head-shadowing effect, which is specific to each individuals head. Therefore, ILD data is often taken from KEMAR studies where two microphones are placed within a fake mannequin head, similar to binaural recordings, and the ILD's are recorded at various angles.

For this paper, the ILD data came from an experiment by Michigan State University, where the level differences were calculated by running simulations of sound waves around idealised models of the human head and torso [3].

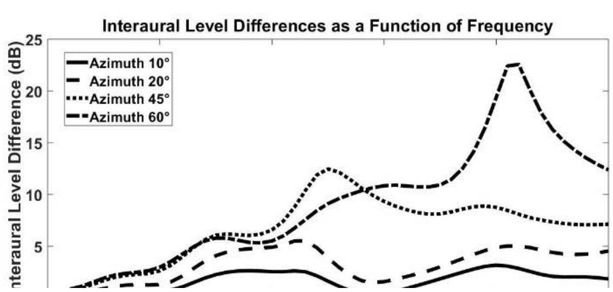

Figure 1- Simulated ILD data as a function of frequency.

The ILD data for 100Hz-2500Hz at angles of 10°, 20°, 30°, 45°, 60° and 90° were simulated and plotted in figure 1. The data for an azimuthal angle of 30° and 90° was omitted to make the figure less cluttered. This simulated data was chosen over experimental data because, whilst it fitted the experimental data collected in KEMAR studies well, the parameters such as head radius were also tunable instead of taken as an average of numerous listeners.

2.3. Pinna Reflections

Whilst the ITD and ILD parameters of a stereo sound account for its azimuthal angle, θ, its angle from the zero point in the horizontal plane. It is the reflections caused by the outer ear (pinna) that give a sound an 'elevation', φ, in the vertical plane [4]. There are two quantities associated with each pinna reflection, a reflection coefficient ρn and a time delay τn. The number of reflections caused by the pinna, n, has been found to be 3 or more for a source in front of the listener or 2 for sources behind the listener. This is if n=1 is considered to be the source hitting the ear directly since rear sources have been found to only have one major reflection [5]. τn is given by the equation:

$$\tau n(\theta, \emptyset) = A_n \cos\left(\frac{\theta}{2}\right) \sin\left(D_n(90° - \varphi)\right) + B_n \tag{3}$$

Where A_n is an amplitude, B_n is an offset and D_n is a scaling factor. The equation, coefficients and values for τn were all found from research by Brown and Duda [6]. The coefficients are as follows:

n	ρ_n	A_n (samples)	B_n (samples)	D_n
2	0.5	1	2	1
3	-1	5	4	0.5
4	0.5	5	7	0.5
5	-0.25	5	11	0.5
6	0.25	5	13	0.5

Table 1- Table of coefficients for pinna delay.

The coefficient D_n was changed per listener from 1 and 0.5 to 0.85 and 0.35 respectively [7].

3. MATLAB APPLICATION

MATLAB was used to create a standalone application that could take a monoaural input and apply the mian functions of duplex theory described in the previous section to give the sound a perceived azimuthal angle and angle of elevation. The sound would then be outputted to the left and right channels of earphones as a stereo sound. Since the pinna reflections are simulated within the app, only in-ear earphones would be appropriate as they bypass the listeners outer-ear.

Once a method for calculating the ILD, ITD and pinna reflections the order of which each of these effects had to be added to the monaural sound had to be decided. Although pinna reflections are affected by the azimuthal angle, this has been proven to be mostly undetectable by listeners [6]. Therefore, the pinna reflections could be applied to the monaural input before it is split to the left and right channels. A flow chart of the code can be seen below.

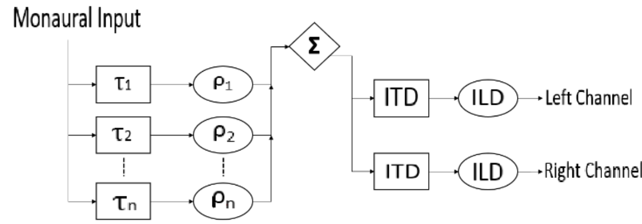

Figure 2- A flowchart demonstrating the order that each transformation was applied to the audio signal.

The Digital Audio Effects handbook [8] was used as a guide when creating the required effects within the app as well the app itself. Most of the functions used within MATLAB were taken from the audio toolbox available with MATLAB19. The initial sound was imported to MATLAB and stored in a one-column array, to ensure the sound was monoaural.

The variables for azimuthal and elevation angle are inputted to the app using sliders, seen in figure 3. The resulting binaural audio file can be listened to using the test button and then saved as a .wav file. The completed application can be downloaded from https://drive.google.com/drive/folders/1YpfQ7eAUzTc7Z5aUX8w9P7jS1THTtZ21?usp=sharing.

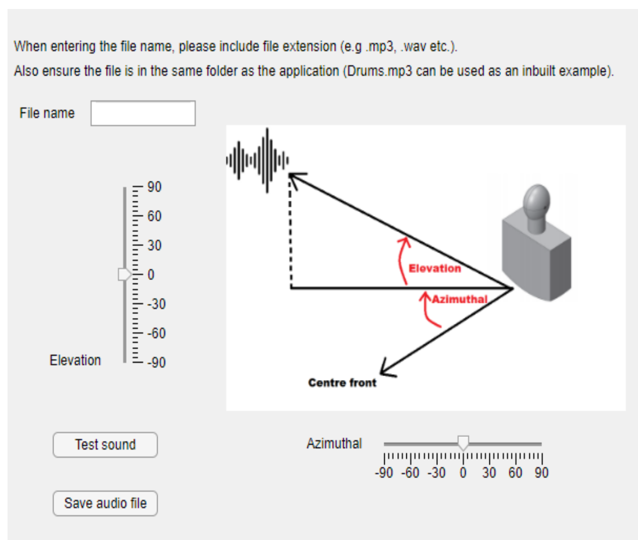

Figure 3-GUI for spectral panner app.

3.1. Delayed Signals

The delay effect needed to apply the pinna reflections was applied using the 'dsp.delay' within MATLAB. Since equation 3 gave a delay in samples no conversion was necessary, each reflection was simply multiplied by the reflection coefficient and added to the original signal. An example of an 800hz sine wave before and after applying the pinna reflections can be seen below.

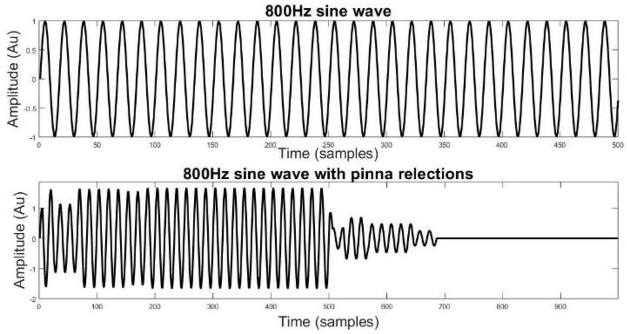

Figure 4- 800Hz sinewave before and after pinna reflections for a 0° elevation are applied.

For the ITD, equation 2 calculates a delay in seconds. This had to then be multiplied by the sample rate (taken from the original audio file) to give a delay in samples. The same ‘dsp.delay’ was then used to delay either the right or left channel.

3.2. Parametric Equalizer

In order to apply the frequency specific ILD within the MATLAB app, a parametric equalizer needed to be created. This was an array of frequencies with an associated gain/reduction that could then be applied to the audio signal as a filter. This was done by using the inbuilt ‘designParamEQ’ within MATLAB and creating an array from the simulated data [3] that could be changed depending on the input azimuthal angle.

An example of the difference between left and right channels once both the ITD and ILD filter had been applied to a recording of drums can be seen in figure 4. The recording was affected by an azimuthal angle of 60° and an elevation of 30°, this has caused a considerable change in the amplitude and small changes noticeable in the waveform from the effects of the ITD and Pinna Reflections.

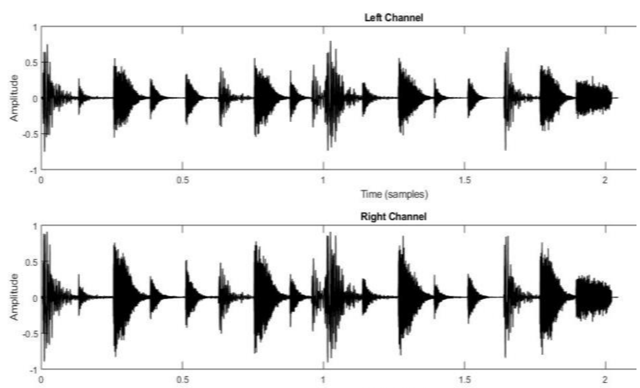

Figure 2-Left and Right channels of a sound recording affected by ITD and ILD.

Once the desired effect has been applied, the two-column array (for a stereo sound) can be written as an audio file that now has the desired spatial quality.

4. CONCLUSION

The research that preceded this paper found that the most common method of replicating spatial audio digitally was by applying a Head Related Transfer Function from an online catalog. This paper has proposed a method of recreating this effect digitally without the need of any experimental data. An application was created that uses a variety of functions to apply the characteristics associated with sound localization within humans through minimal inputs; head radius, azimuthal angle and angle of elevation to any recorded monoaural sound.

Although not thoroughly tested, preliminary tests, completed by asking 4 participants to identify where the sound appeared to be coming from, have suggested that the application not only replicates the require effect well but stands up against methods such as HRTF transforms and binaural recordings.

5. FUTURE WORK

Now that the application has been shown to replicate the intended effect the next step would be to thoroughly test the application with a variety of inputting sound files and numerous azimuthal and elevation angles. Though it is hard to characterize the effectiveness of a spatial audio application, due each person’s specific perception, it is the intention of the author to test the application with a sample group. A similar approach to that seen in (Brown, C.P 1998) would be employed to test the applications effectiveness.

Once the reliability of the application was confirmed by testing the next step would be to explore the numerous applications. Using MATLAB’s vst generator within the audio toolbox to create a virtual effect for use within Digital Audio Workstations (DAWs) would be something that could explored. This would create a more ‘powerful’ panning tool than the typical inbuilt ones seen in most DAWs that simply reduces the amplitude of each channel without being as over-complicated as some of the more complex ‘spatial audio’ vst’s on the market.

Another exciting application would be to stream azimuthal and elevation angle data to MATLAB from a wireless device and account for head movements within the software. This would provide the basis for effective tracking/guidance software.

6. REFERENCES

(1) Rumsey, F. (2001). Spatial Audio. Oxford: CRC Press.
(2) Duda, R.O (1993) 'Modeling head related transfer functions', Proceedings of 27th Asilomar Conference on Signals, Systems and Computers, (), pp. [Online]. Available at: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=342420 (Accessed: 10/03/19).
(3) Tingli, C. Rakerd, B. (2015) 'Computing interaural differences through finite element modeling of idealized human heads', The Journal of the Acoustical Society of America, 138(3), pp. 1549–1560 [Online]. Available at: (Accessed: 10/03/19).

(4) Wright, D. Hebrank, J. (1974) 'Pinna reflections as cues for localization', The Journal of the Acoustical Society of America, 56(957).

(5) Hiranaka, Y. Yamasaki, H. (1983) 'Envelope representations of pinna impulse responses relating to three-dimensional localization of sound sources', The Journal of the Acoustical Society of America, 73(291).

(6) Brown, C. P., Duda, R.O. (1998) 'A Structural Model for Binaural Sound Synthesis', EEE Transactions on Speech and Audio Processing, 6(5), pp. 477 [Online]. Available at: (Accessed: 2/04/19).

(7) Brown, C.P. (1996) Modelling the Elevation Characteristics of the Head-Related Impulse Response, San Jose State University

(8) Zolzer, U. (2011). DAFX. Chichester, West Sussex, England: Wiley.