

UNIVERSITY OF EDINBURGH

School of Physics and Astronomy

Binaural Audio Project

Roberto Becerra

MSc Acoustics and Music
Technology

S1034048

s1034048@sms.ed.ac.uk

17 March 11

ABSTRACT

The aim of this project is to expand on the techniques and knowledge used in binaural audio. This includes main characteristics: Interaural Time Difference (ITD), Interaural Level Difference (ILD) and Head Related Transfer Function (HRTF). Recordings were made at the University's anechoic chamber with a dummy head and binaural microphones to test the effect of turning the head in front of a speaker. The recordings done included a range of pure tones at different frequencies, white noise and sine sweeps. Programs were done in MATLAB to determine ITDs and IILs as well as HRTFs based on Fourier analysis and cross correlation and autocorrelation of the sounds recorded at the microphones and the sounds played. The outcome of the project was a set of binaural cues and data used to generate transfer functions that can be applied to *dry* mono sounds to perform virtual localization on them.

Declaration

I declare that that this project and report is my own work.

Signature:

Date:

Supervisor: Clive Greated

6 Weeks

Contents

ABSTRACT.....	1
NOTATION.....	3
1 INTRODUCTION.....	3
2 LITERATURE REVIEW & BACKGROUND.....	4
2.1 SOUND PERCEPTION.....	4
2.2 SOUND SPATIALIZATION.....	5
2.2.1 INTERAURAL TIME DIFFERENCE (ITD).....	7
2.2.2 INTERAURAL LEVEL DIFFERENCE (ILD).....	8
2.2.3 HEAD RELATED TRANSFER FUNCTION (HRTF).....	9
2.3 MATHEMATICAL TOOLS USED TO DETERMINE EXPERIMENTS DATA.....	9
2.3.1 FOURIER THEORY.....	9
2.3.2 CONVOLUTION.....	10
2.3.3 CROSS CORRELATION.....	10
3 PROCEDURE.....	12
3.1 OVERVIEW.....	12
3.2 CONFIGURATION.....	12
3.3 MEASUREMENTS.....	13
3.4 ILD, ITD AND TRANSFER FUNCTIONS DETERMINATION.....	13
4 RESULTS AND DISCUSSION.....	19
4.1 INTERAURAL TIME DIFFERENCES.....	19
4.2 INTERAURAL LEVEL DIFFERENCES.....	20
4.3 HEAD RELATED TRANSFER FUNCTIONS.....	21
5 CONCLUSION.....	22
6 REFERENCES.....	23
7 TABLE OF FIGURES.....	24
8 TABLE OF EQUATIONS.....	24
A APPENDICES.....	25
INTERAURAL TIME DIFFERENCES DATA.....	25
INTERAURAL LEVEL DIFFERENCES DATA.....	25

NOTATION

Along the text the abbreviations IID and ILD can be used for the same concept, as they describe different names this cue has been given by people in different times. The same happens with ITD and IPD to describe time or phase differences.

1 INTRODUCTION

In space sound sources are localized by our brain using different interaural (referred to both ears) cues. The brain interprets the difference in time and amplitude on the sound arriving to each ear, i.e. if the sound signal arrives with a phase difference first to the right ear and then to the left ear, we know the sound is coming from the right hand side. The time difference and resonances are due to the shape of the body, head and pinna as well as the loudness yield to a perception of the position and distance of the perceived sound source. Other physical factors are involved in the perception of human hearing, making it a highly complex system.

This theory was first expressed by John Williams Strutt, third Baron Rayleigh, at around 1900, (1) sound sources are localized by perception of Interaural Time Differences (ITD) and Interaural Level Differences (ILD). This theory is now known as Duplex Theory.

Sound arriving at the ears is frequency altered or filtered according to the physiognomy of the listener; this includes the shape of the head, ear and torso. This filtering effect is known as Head Related Transfer Function (HRTF), and it is of major importance in the localization of sound by the brain as in the synthesis of virtual spatialized audio.

It is possible to simulate these cues and localize a “dry” sound by reproducing it through a headset. Several applications of these techniques are being developed and research is being done on the subject, such as didactic interactions in museums, augmented reality installations, and using it to develop much more complex systems like Wave Field Synthesis (WFS) (1) (2).

This project aims to acquire empirical HRTFs and data related to binaural audio and explain the basic knowledge around this theory. This data will be then used to synthesize sound samples with virtual 3D localization that can be listened to through headsets.

2 LITERATURE REVIEW & BACKGROUND

Related work on this subject ranges from the early studies done to develop ITD and ILD, to novel applications on the binaural techniques done mostly in research parties such as MIT, IRCAM and IMK. (1) (3) (4) Both techniques use headphones and stereo speakers.

Long ago Lord Rayleigh (5) worked on the cues that let the auditory system determine where a sound source is located on the environment in which humans wander. His work and statements are widely referenced as they first introduced the concepts of ITD and ILD. He first compared the auditory ability to find a sound source within an auditory field with the visual skills performed to do the same with visual stimuli. Although his work is focused on pure tones it has a great influence on the work done these days.

Work is also done on therapeutic, behavioral psychoacoustics applications of the theory of sound; using the physical aspects of binaural audio, it is known that applying two slightly different frequencies on each ear will produce a beating on the perception effectuated by the brain. This beat can range from a few Hertz up to around 30 Hz, inducing different effects on the mood of the listener. Such low frequency beats elicit an entrainment electroencephalogram frequencies and changes on the listener state of consciousness (6).

Efforts have been made to accurately render 3D sound using HRTFs. Measurements are done in anechoic chambers with dummy heads on a dense grid of mapping data to analyze the hearing frequency and time response. Methods using interpolation are used to construct HRTFs values of spots that haven't been mapped, by simply truncating information or using approximate formulas to simulate the non-linear behaviour of the system. This frequency response data is later used to produce filtering effects that mimic these transfer functions (4), (7).

Finally, much more complex analysis techniques have been made to use spherical harmonics calculations and dissect even more the HRTF behavior due to highly non linear aspects of the phenomenon [SPHERICAL HARMONICS], and interesting new render techniques are being used to take binaural aspects of hearing and create novel sound experiences, such as the CARROUSO project, using WFS (2).

2.1 SOUND PERCEPTION

Sound is perceived by the ears by sensing the air pressure of the audio signals coming into them. This pressure then moves the ear drum and this in turn produces movement on the Ossicles bones on the middle ear. These bones then transfer vibration to the liquid media in which the coiled cochlea translates the time domain information into frequency domain electrical impulses that are fed into brain cells (8).

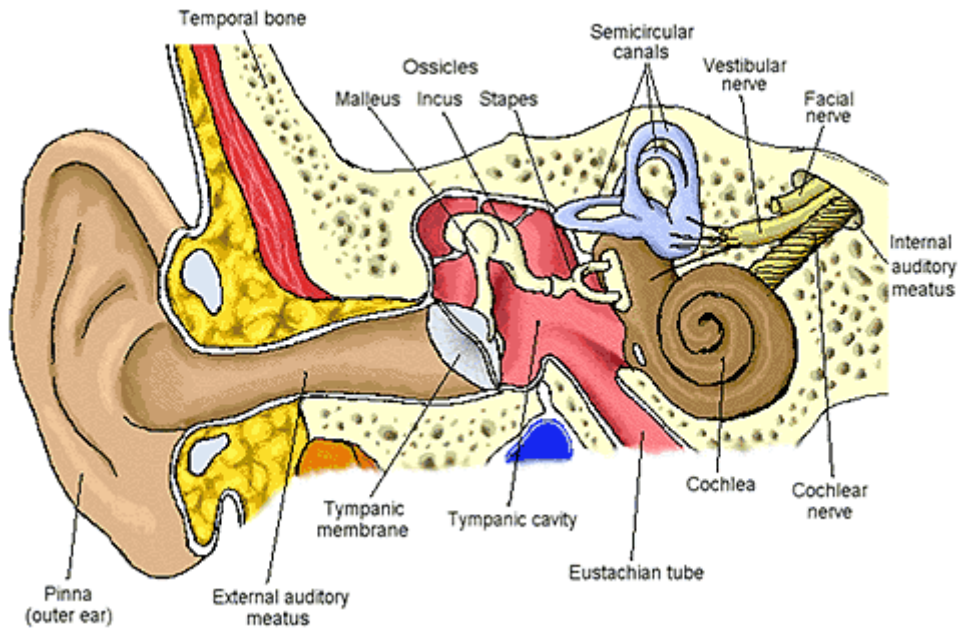


Figure 1 - Human Ear (9)

Human hearing response is limited from 20 hertz on the hearing threshold, up to 20 kHz on the limit of high frequencies that can be perceived. Though there is more to say about how the hearing organ reacts to sound pressure, as measurements have been done to show that some frequencies are more easily perceived than others. This yields to a naming of the scales used to measure sound pressure and sound loudness as perceived by humans; Sound Pressure Level (SPL) is normally scaled in dB with respect to the lowest pressure that can be heard at 1000 Hz, and Loudness level is a physical term used to describe the loudness of a sound and its unit is Phon (10).

Phones express the circumstances in which human ears perceive a sound of different frequency or SPL as being of the same loudness. In Figure 2 the Equal Loudness Contours of Human Hearing is depicted, showing that at low frequencies, a higher pressure level is needed to make the sound perceivable, whereas at higher frequencies, lower pressure level is needed. In fact we can observe that the perception is more sensitive around frequencies of 3 kHz, this is due to the effect known as Helmholtz resonator, in which a certain frequency is busted on a tunnel like structure (ear canal) in function of its diameter and length (11).

This shows that the frequency response on the hearing system is non linear and that non linear events will occur when analyzing sounds coming from different bandwidths and locations on the space surrounding the listener.

2.2 SOUND SPATIALIZATION

Different cues help the brain interpret sounds coming from a real environment in which listeners are immersed. As mentioned before, the most important ones are those that describe the time and intensity differences with which sound arrives at both ears. In addition, a number of cues are of use

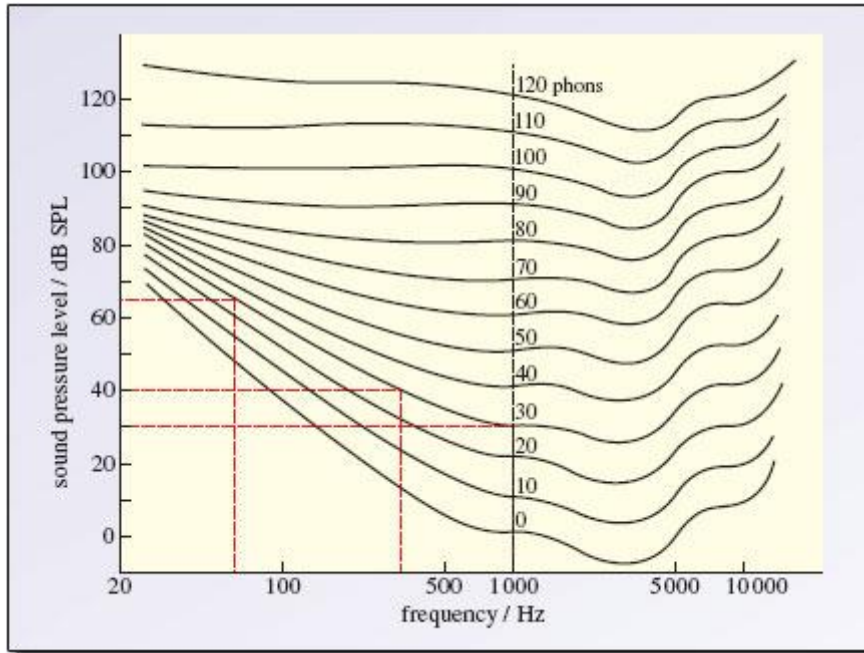


Figure 2- Equal-Loudness Contours

in the aid of human sound localization, such as head shadow, pinna response, shoulder echo, head motion, early echo response/reverberation, and vision. The first three cues are considered to be static, whereas the later are referred to as dynamic cues (12). As mentioned, the response to the sounds approaching the head can be thought of as a sophisticated system with various factors involved in the spatialization of sound.

To measure the possible positions from which sounds may approach the head, a number of terms are used that describe the position in space of both sound and listener. The Head Related Coordinate System is shown in the figure below; this is useful as we need to express different positions that produce different filtering effects, ITDs and ILDs.

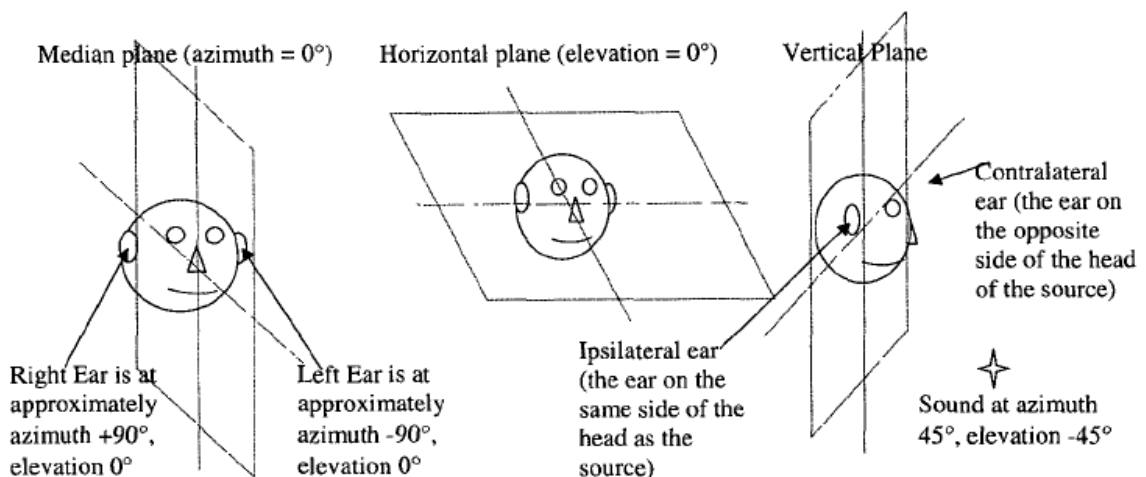


Figure 3- Head Related Coordinate System (13).

For the purpose of this project only variations on the horizontal plane were measured, meaning that the dummy head that was used and mounted on a turning table was turned on this axis, causing the

sound to come from different angles in this plane, in this case, this angle measure is called Azimuth; so for the occasion in which the head is facing the speaker from which the sound is coming, an Azimuth of 0° is considered; when the head's right ear is facing the speaker, an Azimuth of 90° is defined, and so on.

2.2.1 INTERAURAL TIME DIFFERENCE (ITD)

Air pressure travels across the air and takes some time to arrive at each ear. This time is given by the speed of sound and the distance sound has to travel to either ear. For instance if the sound is on the right hand side of the listener, the sound pressure will arrive first at the right ear and after a few moments it will also reach the left ear; sound had to travel around the head. Thus a time or phase difference is accounted as a significant cue for the brain to know where the sound is coming from.

Simple trigonometric calculations can be done to estimate the additional distance sound has to travel to reach the farthest ear on an auditory event. For this purpose the head is modeled and simplified as a sphere with constant diameter. It is clear that this approximation will distort the actual effect of the sound around the head and will give inaccuracies on estimations, but these are taken to be very small and not determinant for the purposes of these studies. Figure 3 shows the modeled shape of the head with the parameters taken into count to calculate the additional time or phase added to the ear receiving sound pressure at last.

It is important to notice that these ITDs are useful only to a certain degree, as different wavelengths yield to alias interpretation problems, because only frequencies which ITD s are only half the period of waveform of that frequency (14). This means that tones above 1500Hz are not interpreted correctly as the waveform repetitions yield to confusions on which part of it arrived first. At this point, when working with pure sinusoidal waves, the term phase difference (IPD) is used alternatively with ITD, as they express the same effect.

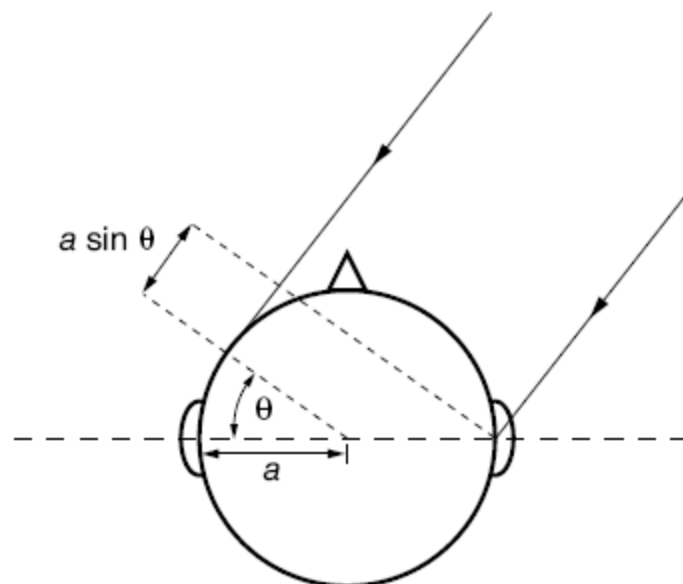


Figure 4 - Modelled head with parameters to measure time difference (14).

Then, if the head is modeled as the picture above, calculations on the extra distance sound has to travel, meaning the ITD of it coming from θ can be computed with the following expression

$$\tau = \left(\frac{a}{c}\right) 2 \sin \theta$$

Equation 1

for frequencies below 500 Hz and

$$\tau = \left(\frac{a}{c}\right) (\theta + \sin \theta)$$

Equation 2

for frequencies above 2 kHz, being a the radius of the head (approximately 87.5 cm) and c the speed of sound (14).

ITD being an inefficient cue for determining the sound localization of a source is no longer a problem when a broadband and/or non periodic sound is played to the listener because it carries more information that can be used to describe where the source is on space.

2.2.2 INTERAURAL LEVEL DIFFERENCE (ILD)

Together with ITDs, ILDs are vital to locate a given sound source in space, because it is a cue which tells us how far an object is or which ear is closer to that source. In opposition to the cue treated above, ILD is more efficient at higher frequencies, as their loudness is altered by the *shadowing effect* of the head. In other words, when a sound has a low frequency component, the air pressure is well distributed around the head, because this is smaller or shorter than the sound's wavelength, whereas the head interferes with high frequency sound wavelengths and their movement towards the farthest ear of the listener, as depicted on the next figure.

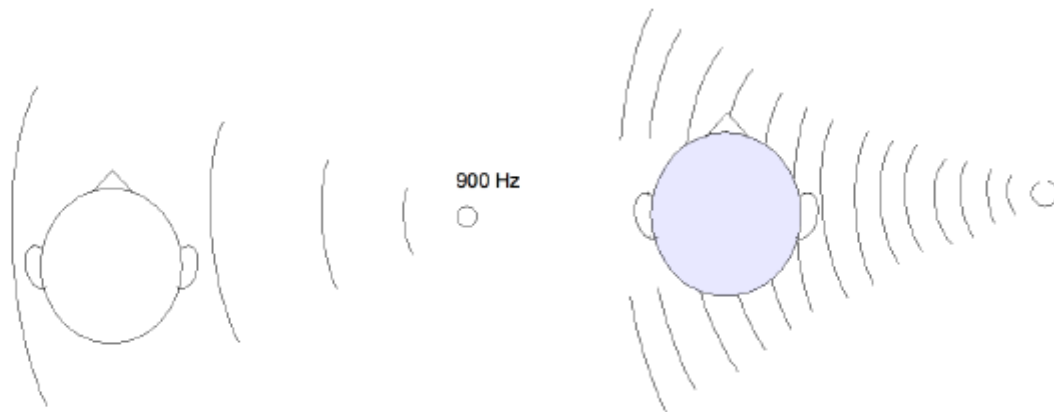


Figure 5 - Waveforms of low and high frequencies, and shadowing effect (4).

So for pure tones at low frequencies it is hard to tell where the sound is coming from, as the level difference is quite small, whereas above the threshold in which the wavelength starts to suffer from the size of the head, this shadowing effect becomes more evident, causing more notable level differences between ears.

Again, when a broadband sound is used it is easier for the auditory system to break it into bits that are useful for ILD analysis that works for localizing the sound source.

These two interaural differences working together give us very useful information about the distance, and origin of the sound source we are listening to, but still when synthesizing 3D audio

they are not sufficient to create an appropriate virtual auditory illusion; it takes more modifications on the sound as stated in the next section.

2.2.3 HEAD RELATED TRANSFER FUNCTION (HRTF)

If only ITD and ILD are used to synthesize a sound a lateralization rather than a virtual spatialization is obtained, as the sound perceived through headsets appears to move just from one ear to the other within the listener's head (14). This means that there is still more to know in order to create a credible auditory illusion. This is accomplished by the introduction of the filtering and reverberant effect of the listener's physiognomy on the sound that arrives at either ear.

At this point it can be thought of a frequency response effect of the body on the sound, that depends on the direction the sound is coming from and once again it is clear that the broader the frequency spectrum is, the richer the localization will be. So a frequency domain filter fits in here to simulate the effect of this physiognomy. These types of filters are called Head Related Transfer Functions (HRTFs) and they contain information of how frequencies approaching either ear on this type of system are affected and suppressed/busted. They are called transfer functions as they are such, a division of an output by an input.

A transfer function is defined as a ratio that shows the proportion of the input and output of a system, so if the measurements taken at the ears of the dummy head are divided by the input in the frequency domain we can obtain a HRTF that will include information of the deviations due to the bodily influence on the sound. This can be defined as

$$H = Y \div X$$

Equation 3

where Y and X are the frequency domain expression of the output and input of the system respectively. Then H can also be defined as expressed by a minimum phase and an allpass system (7), which can be written as

$$H(z) = H_{min}(z)H_{ap}(z)$$

Equation 4

where H_{min} is a minimum phase system and H_{ap} an allpass system that describe both the magnitude and phase alterations expressed by the transfer function.

2.3 MATHEMATICAL TOOLS USED TO DETERMINE EXPERIMENTS DATA

2.3.1 FOURIER THEORY

Fourier frequency analysis was first developed by Jean Baptiste Joseph Fourier (21 March 1768 – 16 May 1830), a French mathematician. This theory has majorly revolutionized the fields of science and engineering as it allows breaking any signal into its frequency components, thus *forgetting* about the time domain and bringing new possibilities of wave modification by modifying what frequencies exist on a given waveform. The theory is rather large and will not be treated fully here, further references about it can be found at (15).

What has to be clear about Fourier Theory is that any given waveform can be expressed as the sum of pure sinusoids, that when combined result on the signal we analyze. On digital signals this analysis is limited by the sampling rate because the maximum frequency that can be detected on a sample is half the sampling rates, this is due to the Nyquist theorem (15). This analysis is expressed by the following equation

$$X_k = \sum_{n=0}^{N-1} x(n)e^{-\frac{2\pi jkn}{N}}, \quad k = 0, 1, \dots, N - 1$$

Equation 5

for discrete samples. Where X is the frequency domain of x , and N is the length of the sample. This is called the Discrete Fourier Transform. Though this analysis is very useful, it can be very time and resources consuming, so a faster way is majorly used in modern computations, it is called Fast Fourier Transform (FFT). For it to work properly, N has to be chosen as a power of 2.

The FFT is already built in Matlab so it is very simple to get the frequency components of any given signal, although care must be taken to make this signal of length equal to some power of 2, this can be done by appending zeros at the end of the waveform. Once the frequency spectrum is know, modifications or comparison can be done with it, such as HRTFs.

2.3.2 CONVOLUTION

Convolution is a mathematical procedure that is typically used to shape a waveform g in terms of another waveform h so these two functions have two different meanings. Generally, the first is a time dependant signal that is then convolved or *shaped* by the later signal which can be thought of as an impulse response. This impulse response can be that of a room, a speaker, or a microphone on a dummy head (16).

Again this procedure has definition on both continuous and discrete time. For our aims, discrete convolution is used to take a dry sound and give it the characteristics of the impulse or frequency response of the recordings made on either ear of the dummy head. So, discrete convolution can be expressed as the frequency domain product of g and h

$$g * h = G(f)H(f)$$

Equation 6

where G is the Fourier transform of the signal g and H is for the impulse response h . This simplifies things as we can obtain the FFT of any given signal relatively easy.

2.3.3 CROSS CORRELATION

Cross Correlation, is a mathematical procedure closely related to convolution. It is a function that is generally used to know the *correlation* between two signals, by comparing them (17). Correlation of $g(t)$ and $h(t)$ is expressed as $Corr(g,h)(t)$; it exists on the time domain and it is a function of *lag* t . This correlation will grow bigger as the signals are increasingly similar.

As the previous procedures, correlation can be worked on the time and frequency domain, and for this later form it is expressed as

$$\text{Corr}(g, h)_j = G_k H_k^*$$

Equation 7

where g and h are time domain functions and G and H their Fourier Transforms respectively. Notice how complex conjugate of the second function is used, this will affect the sign of the output as follows: if g lags h , i.e. it is phase shifted to the right, the correlation will grow more positive, and vice versa.

When a correlation is performed on the same signal, typically out of phase with itself, the operation is called Autocorrelation and it is a good measure of how shifted the signal is, as the autocorrelation will peak at the moment in time equal to the amount of time the signal is behind itself. This can be successfully used when compared the time differences in the recordings of left and right channel, as in theory it is the same signal presented with a phase difference.

3 PROCEDURE

3.1 OVERVIEW

For this project measurements were taken at the Anechoic Chamber of the University using a dummy head, placed on one side of the room, a single speaker facing the head and binaural microphones to record the response of the different auditory events on several configurations of the head turning. The goal of these configurations was to corroborate and extract ITDs, ILDs and HRTFs in order to create adequate filtering and cues to synthesize localized sounds from dry mono wav files.

Common measurements techniques of frequency response were used as well as software available at the University's computers and labs to analyze the data obtained during these recordings.

3.2 CONFIGURATION

As stated before, simple available equipment was used for the procedures of this project. The recordings were made by playing pure tones, white noise and sine sweeps created on the computer, using Matlab, wave (Farnell Function Generator FG3) and white noise generators (Quan-Tech Noise Generator 420), through a common speaker of commercial frequency response and taken with binaural microphones mounted on either ear of a dummy head. Signals from the microphone were premixed at a hardware mixer (Mackie 1202-VLZ PRO 12-Channel Mic/Line Mixer) to ensure that their gains were equal when the head was positioned with 0° Azimuth. A diagram of the configuration is shown next, followed by a schematic diagram of the whole system.

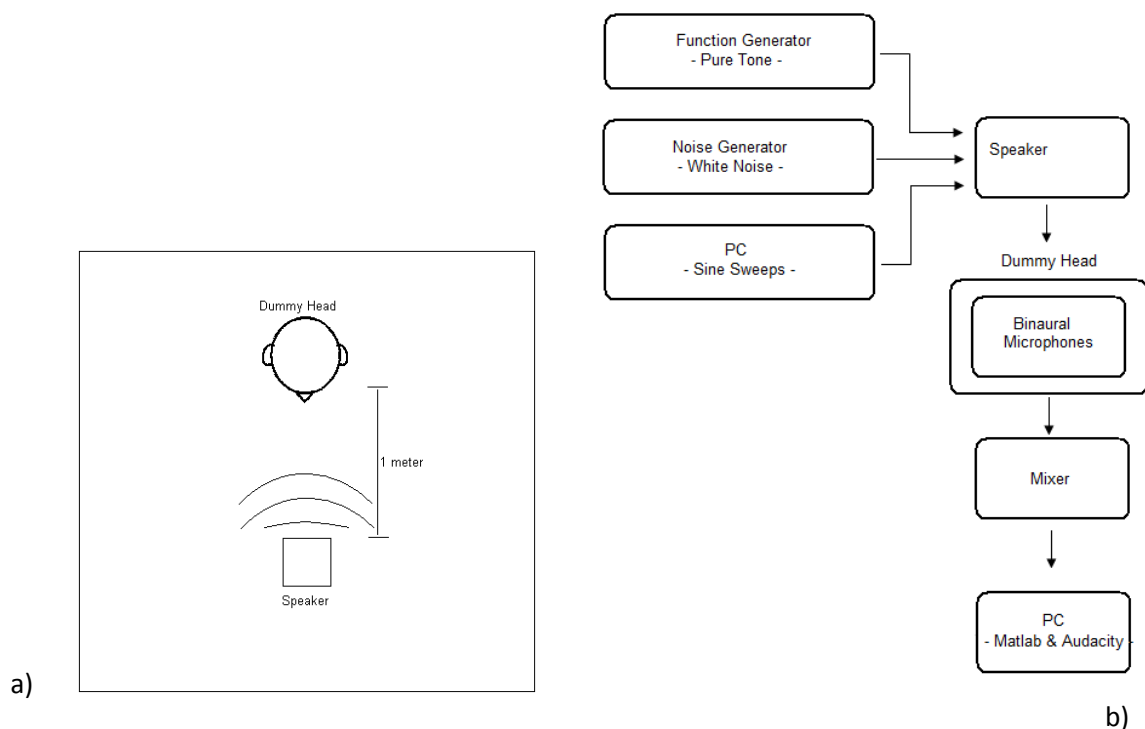


Figure 6 – Configuration and Schematic Diagram. A) Shows how the dummy head and the speaker were arranged at the anechoic chamber and b) Depicts the diagram of the system used for the measurements and analysis throughout this project.

The sound was then analyzed on a PC running Windows XP or Vista, as analysis was made on both the University's and a home computer. During this analysis several mathematical procedures were made to determine the phase and level difference on the pure tone recordings and HRTF on the sweeps and white noise ones.

3.3 MEASUREMENTS

The dummy head was placed on a turning table with a scale to measure the degree by which it was turned on each recording. For all of these recordings, 45° hops were done so 8 points were considered for each case (pure tone, white noise and sine sweep). Audio was played to each of these Azimuth angles for a short period of time, of around 5 seconds, and it was stored on the computer as .wav files that were later chopped in Audacity and analyzed on Matlab.

The first measurements done were with pure tones at 200 Hz, and the intention on this early stage was to extract the phase differences on the recordings at different Azimuth angles and the loudness deviations. These measurements were done to corroborate that the system was being properly designed and to develop the software scripts and methods to do so.

On a later stage of the project more frequencies were included to be recording with the same configuration and Azimuth hops. This time 200 and 300 Hz were recorded to obtain a greater number of measures and validate the previous results.

Moving on, the same type of recordings were made with 200, 300, 1 000, 2 000 and 10 000 Hz and white noise, again keeping the configuration and turning angles. In parallel with these recordings, software was being developed in Matlab to analyze the recordings and so these could be fed into it to be analyzed in an automated way and its results stored on MS Excel spread sheets.

The last stage of the project included same type of recordings with sine sweeps created on Matlab, with a range of frequency from 20 Hz to 20 kHz, corresponding to the hearing range of human auditory system. Again, 8 measurements were done twice and then fed into the Matlab scripts to acquire HRTFs and HRIR to be later convolved with dry sound and create virtual 3D localization.

All of the previous recordings were stored on stereo wav files in which the first channel corresponds to the Left ear and vice versa. Then this files were used in Matlab and split in two different data sets, to separately analyze each channel.

3.4 ILD, ITD AND TRANSFER FUNCTIONS DETERMINATION

Determining of ILDs was made using Matlab programs, in these the wav files were examined to see their dB gain across time so that amplitude could be plotted against time. This was done for both ears and for each one of the frequencies previously mentioned. So at the end a 3D vector was created that shows the changing in amplitude in each ear as the head turns at certain degree intervals.

As expected, level differences were more evident, across a wider range of decibels as the frequency of the measurements increased. So for the testes made for 300 Hz the decibel variance is of nearly 5 dB, whereas for the measurements made at 10 kHz this variance rises up to nearly 30 dB, and if the fact that 3 dB change means double or half of the original loudness, this could indicate that a difference of 1000% from the original loudness is perceived on in both ears.

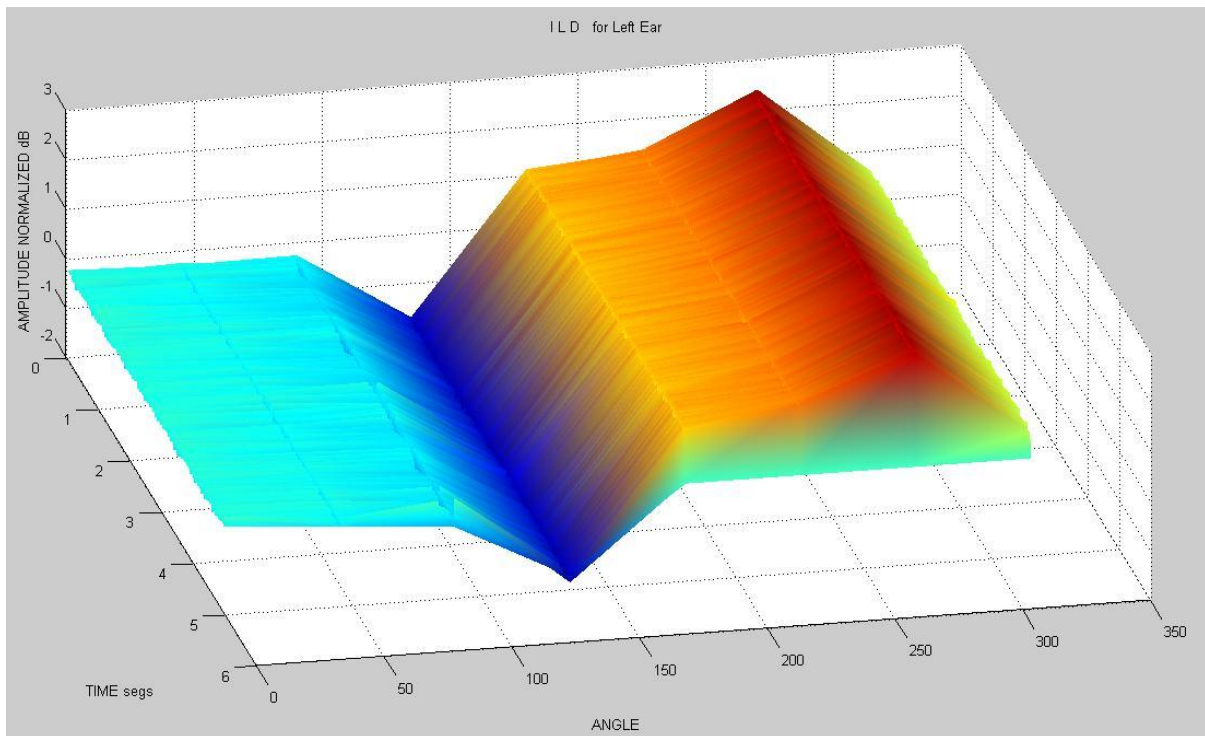


Figure 7 – ILD for Left ear at 300 Hz, it can be observed how the SPL decreases for the first half of the head’s turn and increases on a mirror like behaviour for the next half. Readers should notice the dB range in which this plot stands, as it is of around 5dB, much different from the next figure.

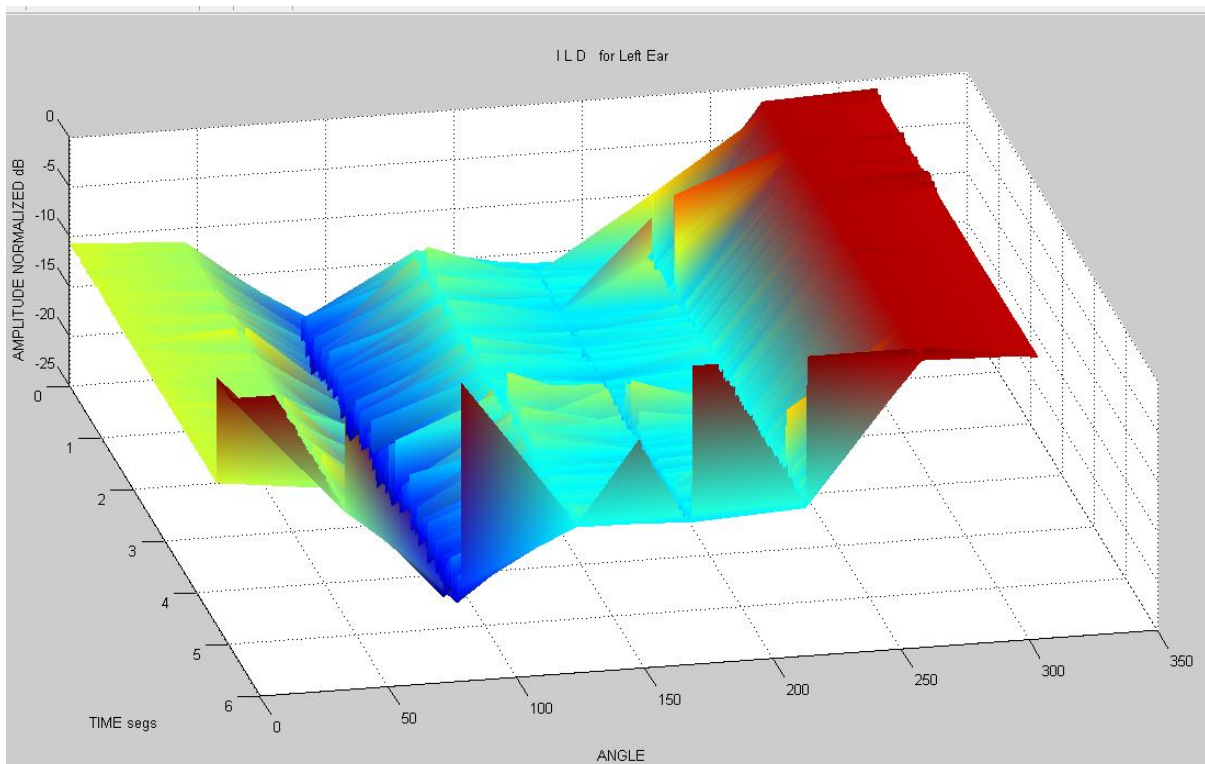


Figure 8 – ILD for Left ear at 10 kHz, readers should notice how the decibels range is much bigger for this plot than that of the previous figure. This is due to the shadowing effect of the head over high frequencies, as the amplitude difference is bigger between the ear facing the sound source and the one shadowed by the head.

Following this measurements, new ones were made to determine the time difference between the signals arriving to the ears. At the end of the day, these are the same signal but with a phase and amplitude difference, so three approaches were applied in this case.

In the first one the signal recorded in each ear was taken and compared with the one from the other ear. From these signals the first period was chopped and compared, looking for maximums and minimums of both waveforms and then extracting the number of samples between them and thus knowing the time difference. This method proved to yield good results, though it had some difficulties due to some noise that was found on the measurements.

The next approach was to take these same two channels' signals and perform a cross correlation between them. This could be referred to as autocorrelation as the two signals are close to be the same signal, delayed from one ear to the other. Because the two signals are periodic and sinusoidal, the output of this correlation was so. So again the maximum value of this correlation was found, as these maximum values indicate how much are the signals related to each other. This max value presumably lies on the time corresponding to the time delay between the correlated signals. Again good results were obtained through this method. To the observer it is notable that in this method it seems like an extra step was added from the previous approach, but in reality an extra step was added by correlating but another step was discarded as just finding the maximum values of one signal.

Despite of this last statement, in practice two correlations were computed, to correlate first the left channel respect to the right and know by how much it was delayed. A later correlation was made of the right channel with respect to the left channel, again to see by how much the left signal was lagging the right one. This was done with the logic that if the turning of the head is done anticlockwise, by the time when the head has turned by 45° the right ear will be getting the sound first, followed by the left ear, and this relationship will be maintained until the head reaches a turn of 180°, when the signal is in phase again. After 180°, it is the left ear that will be getting the sound first. In other words in the first half of the turning, the signal on the left ear lags the one on the right ear, and in the second half of this turning, it is the right ear's signal that lags the left ear's signal.

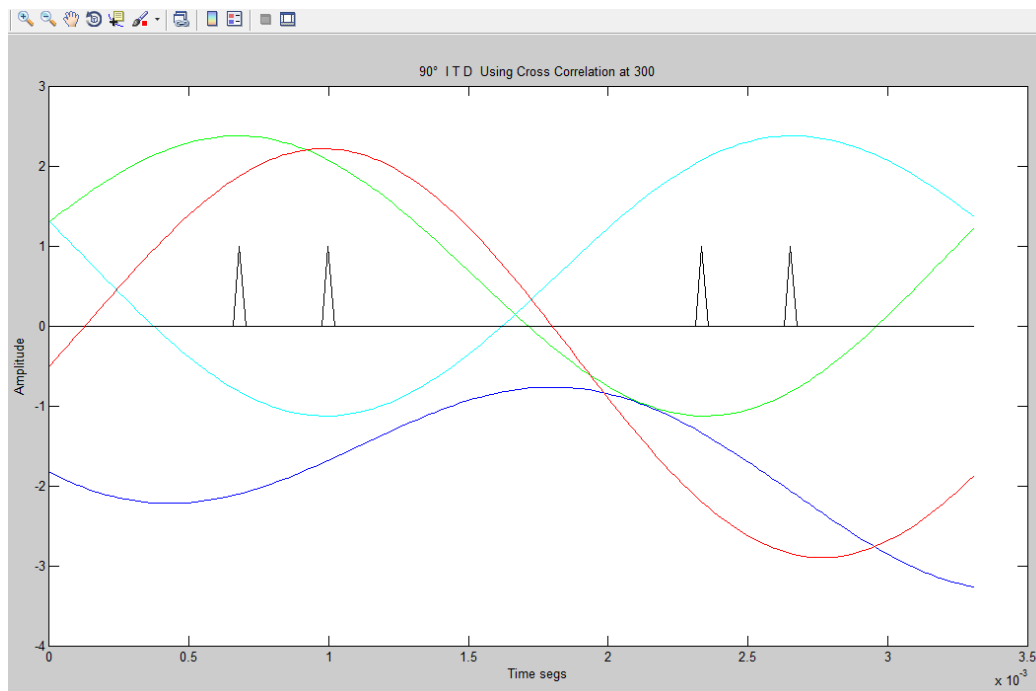


Figure 9 – This figure shows audio from left channel (blue), called l , audio from right channel (red), called r , and two correlation curves. The first one (green) expresses $Corr(l,r)(t) = L(f)H(f)^*$, and the later (cyan) depicts $Corr(r,l)(t) = H(f)L(f)^*$. This is done to get all the max and values of all possibilities: channel 1 lagging channel 2 and the other way around. Each of these maximums and minimums is plotted with a black peak that indicates the time when they happen,

corresponding with the ITD. In this image the recording with the head turned by 90 degrees clockwise places the right ear's audio shifted ahead of the left ears' audio by 8×10^{-4} seconds, as shown by the first peak.

Finally, the third method is somehow mixed with obtaining HRTFs. Again cross correlation was used, but this time to obtain the time difference between the two ears' impulse responses that was measured by playing sine sweeps to the dummy head.

The procedure was to take the signal on each ear and perform a FFT on them, then, to divide this FFT by the FFT of the sine sweep that was produced on Matlab. By doing so a HRTF for each ear channel is obtained, and performing an inverse FFT on this transfer functions results on Head Related Impulse Responses (HRIRs) , that can be later be convolved with any dry sound to produce a spatialization effect. So the correlation was performed between this two HRIR for each of the angles by which the head was turned to get again accurate ITD found on the difference of phase on the HRIRs.

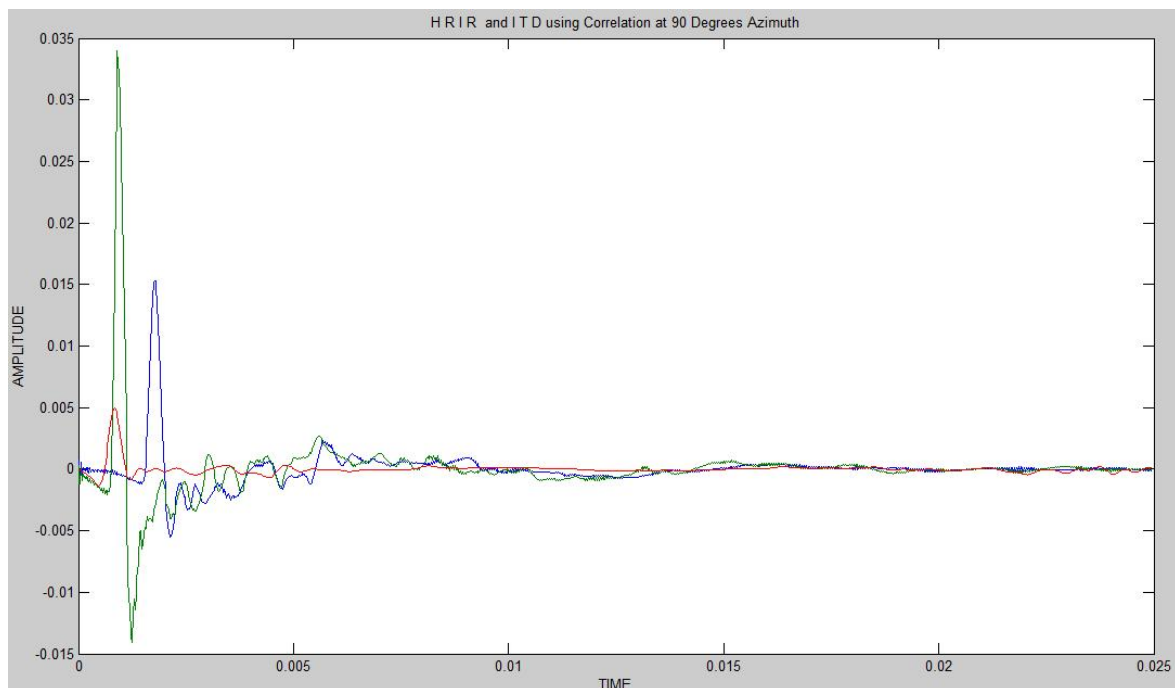


Figure 10 – This image depicts the HRIRs of Left channel (blue), and Right channel (green), plus the Correlation curve (red) generated by doing $Corr(l,r)(t) = L(f)H(f)^*$ with the head turned by 90 degrees and listening to sine sweeps. This curve peaks at time = 8.6168×10^{-4} , which is the same ITD difference that was obtained before, with the methods mentioned above.

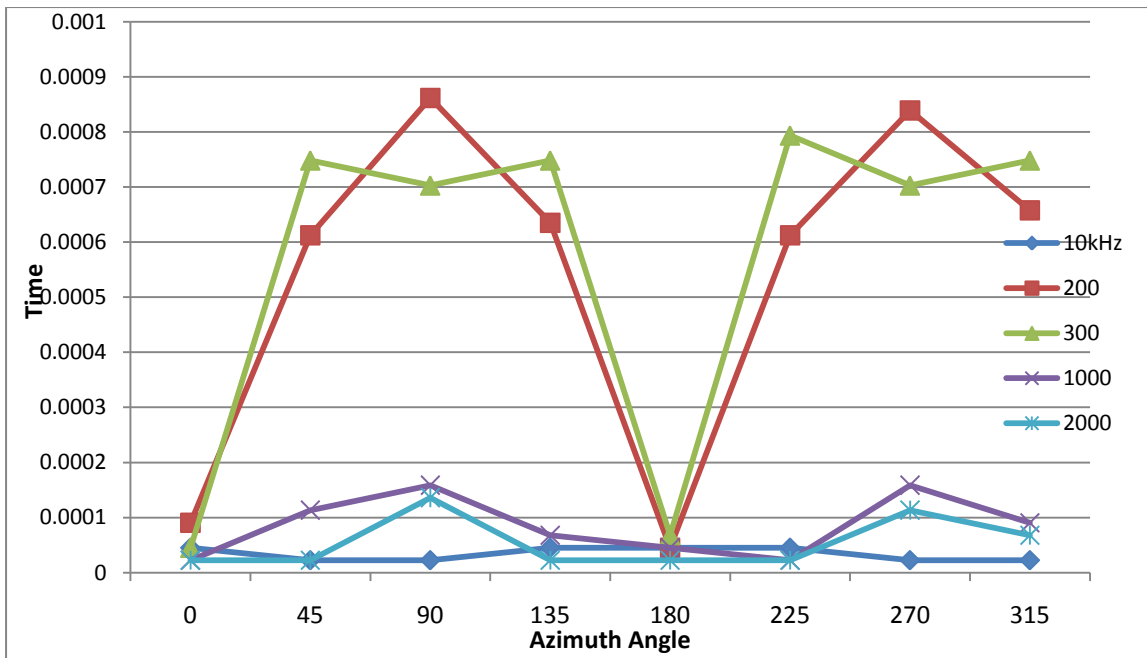


Figure 11 - This plot shows the different ITD obtained for all angles and frequencies measured, with the three methods used and mentioned before.

It is important to mention that the same ITD quantities were obtained with either of the methods described previously, so the data on Figure 8 relates to all of them.

Finally, the acquisition of the HRTFs was performed as it has been briefly mentioned on the paragraphs above, and now is further described.

HRTFs were computed by recording sine sweeps generated on Matlab that went from 20 to 20000 Hz (Human hearing range) over a time of $2^{19} / SR$ seconds, being SR the sample rate, which in this case was 44100. These sweeps were played through the speaker and recorded by the binaural microphones placed on either ear. Then the stereo audio from the dummy head (right and left ear) was split into right and left channel (L and R respectively) and a FFT was performed on each one of them, so a new vectors called L_F and R_F were obtained. A FFT was also applied to the audio vector containing the sine sweep (s), thus creating a new vector called s_F. By dividing L_F by s_F and R_F by s_F, HRTF are computed for each ear.

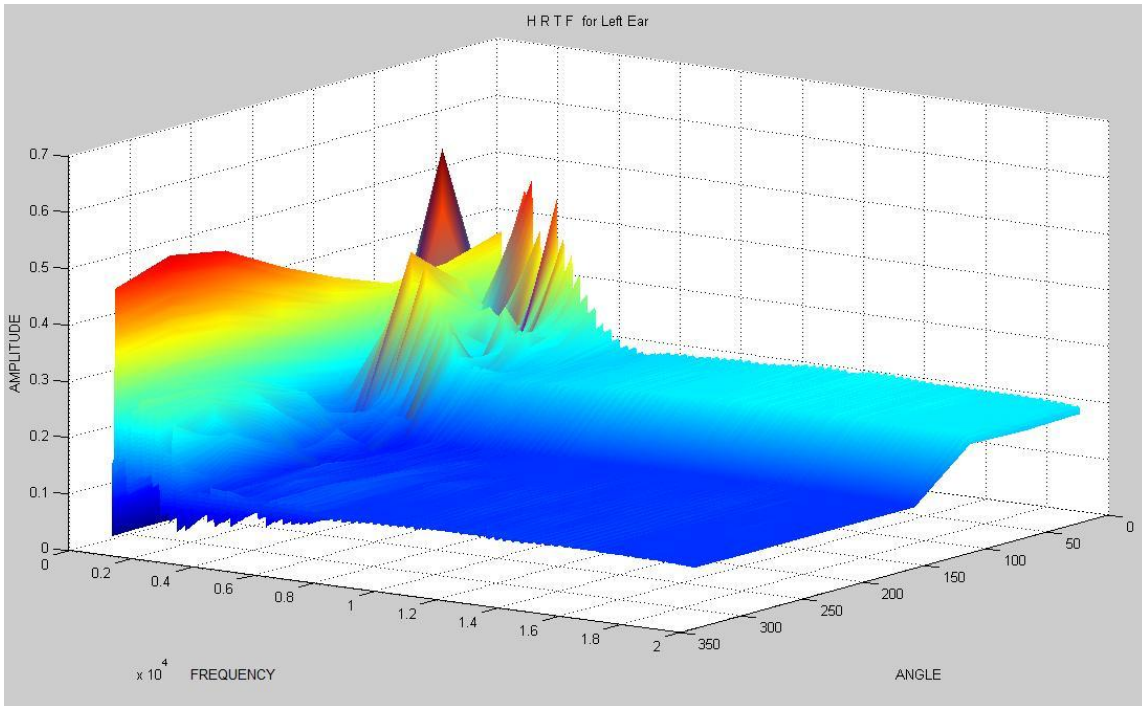


Figure 12 – HRTF of Left ear by dividing $L_F(f)/s_F(f)$, where L_F is the FFT of the left ear audio and s_F is the FFT of the sine sweep. The figure show the behaviour of the HRTF at different turning angles.

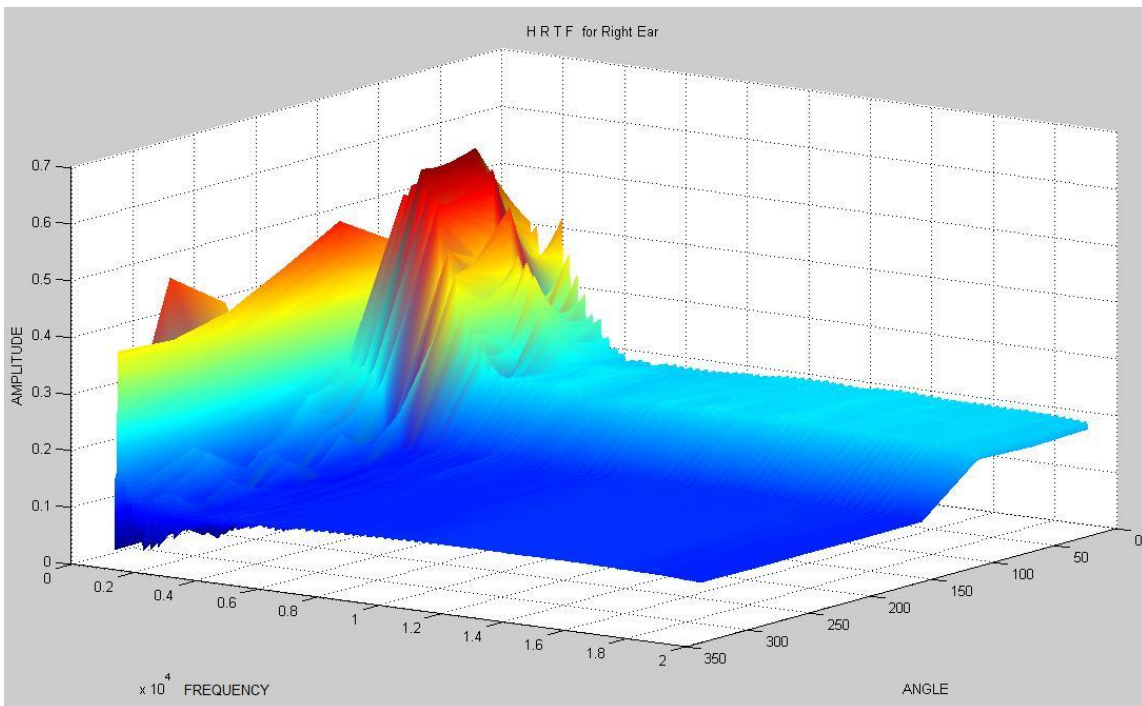


Figure 13 – HRTF of Right ear by dividing $R_F(f)/s_F(f)$, where R_F is the FFT of the right ear audio and s_F is the FFT of the sine sweep. The figure shows the behaviour of the HRTF at different turning angles.

4 RESULTS AND DISCUSSION

4.1 INTERAURAL TIME DIFFERENCES

Correlation was used to determine the time differences between the left and right channels on the dummy head. This mathematical operation come in handy when two similar signals are compared and the aim is to know how or in which measures are they related, and it can be used for (17) extracting phase differences between the same signal, in which case it is called autocorrelation.

After performing such correlation, time differences were determined for all of the frequencies and angles stated before, the next figure shows all the ITD of the two ears for measurements made at 300 Hz. Correlation was used on this case. This figure is actually an expansion of figure as shown in the following figure 6, in which only the case for a turn of 90° was shown.

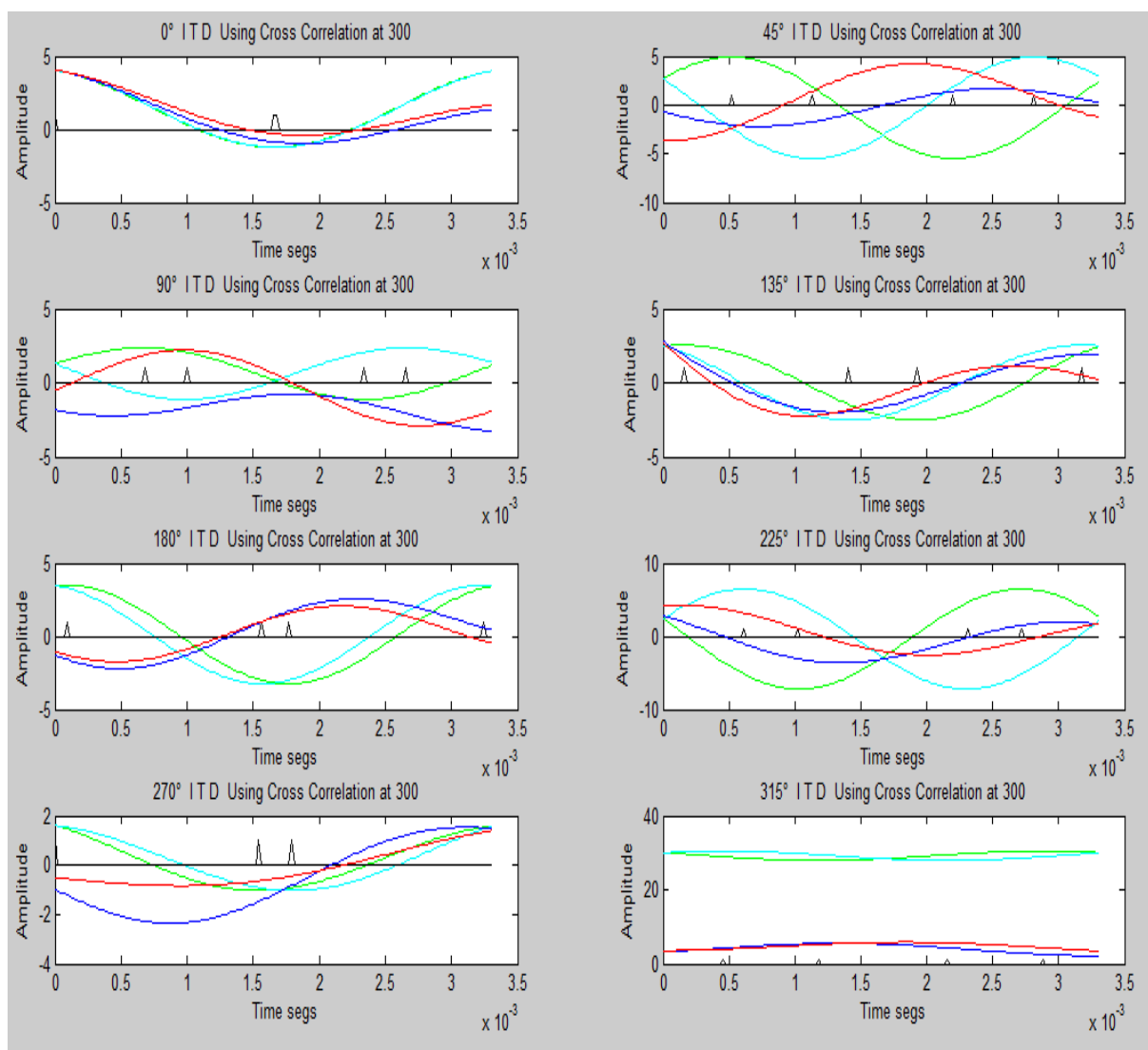


Figure 14 - Correlation and ITDs of low frequency sounds at 300 Hz, in this figure the reader can notice that the subfigure for the ITD at 90° has been shown before. This set of plots shows how the correlation was used to draw peaks at the points in time in which it had maximums –or minimums - (peaked). These peaks stand at the moment in time that represents the ITD between left channel (blue) and the right channel (red). Correlation cuves are show for $Corr(l,r)(t) = L(f)H(f)^*$ (green), and $Corr(r,l)(t) = H(f)L(f)^*$ (cyan).

This picture shows the correlation for the 8 positions described before, starting from top left, moving to the right and then to the next row and so on, and finishing on the bottom right. Blue and red lines show the left and right channels signals, whereas cyan and green lines represent correlation functions show for $Corr(l,r)(t) = L(f)H(f)^*$ (green), and $Corr(r,l)(t) = H(f)L(f)^*$ (cyan). This analysis outputs a sinusoid like correlation line that was observed to indicate where the maximum and minimum occurs. This max and min represents how similar the signals were on the period taken to measure, this means the more similar they were, the bigger the correlation value would be.

It is very important to notice that although several methods were used to compute the ITD on all cases of Azimuth angles and frequencies, the results were always the same. This means that on point to measure is the effectiveness of those methods, how time consuming they are on machine cycles and human work. In the second approach extra unnecessary work was performed by doing two correlations, before knowing one would suffice by doing some extra calculations based on the analyzed period's length. To my experience it seems like the last method was the most effective because it yielded ITDs at the same time it walked forward to obtain HRTFs, so two different objectives are accomplished at once.

4.2 INTERAURAL LEVEL DIFFERENCES

As to ILDs the measurements and analysis are quite simple, it takes a little effort of tuning, which is missing here, so results should be normalized. As expected, level differences vary on a larger scale as the frequency increases; this is due to the shadowing effect of the head over small wavelengths. This is depicted on the next figure, which plots all the results obtained for this project.

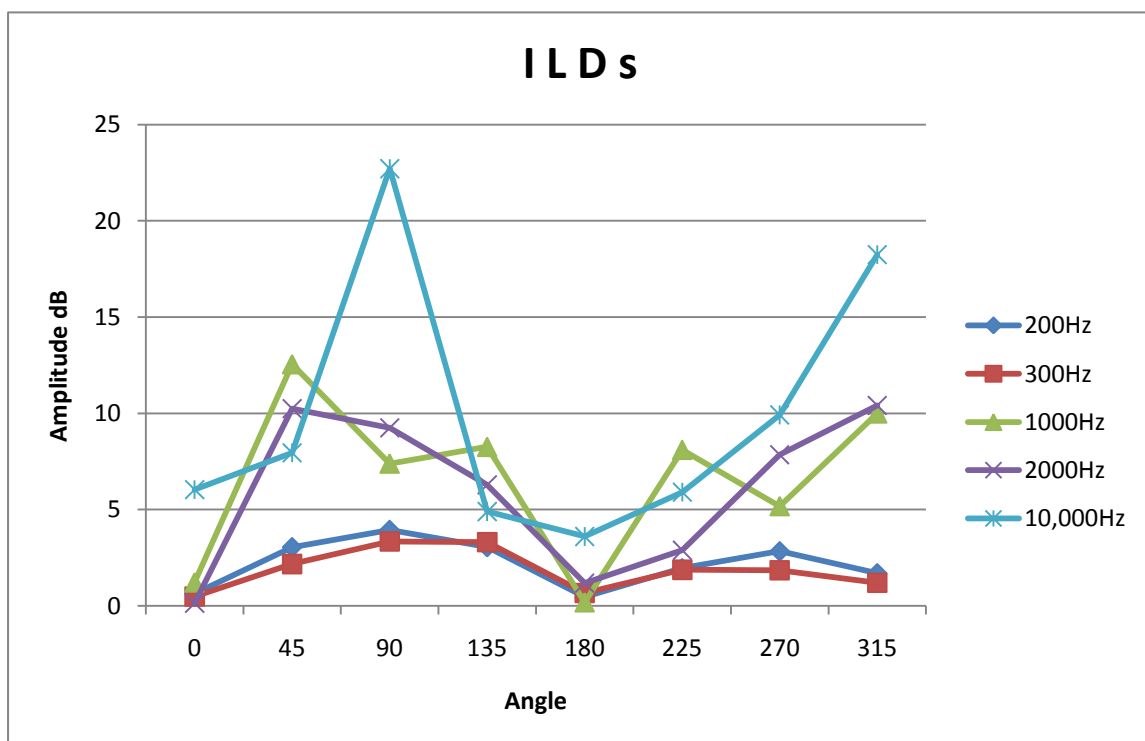


Figure 15 – ILDs obtained for all frequencies and angles measured. This clearly shows how the level differences rises dramatically as the frequency increases and the head turns, due to the shadowing effect of the head.

The current information can easily be used as ratios to synthesize spatialized sounds. Further research can be done on the SPLs for different distances and energies from sound sources.

4.3 HEAD RELATED TRANSFER FUNCTIONS

With the procedure mentioned before, wav files with HRIRs were obtained and thus simple 3D sounds can be synthesized out of them. It can be even a work on interpolation to find ITD, ILD and HRTF's characteristics of points that were not measured.

It can be found on (7) that the magnitude determination of the transfer function frequency domain is of more importance in the convolution with dry sounds to synthesize a localized source, than that of the phase or time differences, as this later can be interpolated from the empirical measurements taken in the field. Thus, the frequency impulse response of the system can be convolved with a given sound and then given the time difference on a different process to simulate this virtual spatialization.

Errors can be found on the sweeps recordings as unintentionally the original sound, coming from the computer was digitally recorded altogether with the perception on the binaural microphones; this produced some inaccuracies, though not so determinant at the end. Another error was done by not knowing of the software capabilities: when recording the sweeps played to the head, a short time was spent between the beginning of the recording and the start of the sweeps; this was due to the manual activation of these functions, while they could have been done automatically and simultaneously. This produced inaccuracies on the HRTFs, causing echoes when convolving dry sounds. To get round this the signals were chopped to get rid of the extra time.

5 CONCLUSION

As to the aim of the project, that was to do research on the different characteristics of Binaural Audio, the results have been satisfactory for a clear understanding of these characteristics and methodologies used to measure and create them. The work done here can result, with further efforts, on the increase of knowledge that can be used to perform more realistic 3D sound renderings. I think that understanding on this field is important to move on to much more sophisticated real time sound rendering techniques, such as WFS as done by (18). And although the methods and sound proposed on this paper are limited to the use of headsets, the concepts of impulse responses, transfer functions and localizing cues is valid for broader applications. Further work can be done on creating a more dense net of measurements, with smaller angle hops and not restricting to just the horizontal plane, but to expand to the Median and Vertical Plane, to create much more richer sets of data. Also work can be done on creating image sources with the inclusion of reverberation on the sound stimulus recorded at the ears of the dummy head. In general, the data and info provided here can be used to further development of better 3D sound synthesis as well as DSP.

6 REFERENCES

1. **Warusfel, Oliver and Eckel, Gerhard.** *LISTEN Augmenting everyday environments through interactive soundscapes.* IRCAM, Fgh-IMK. Paris; Bonn : IRCAM, Fgh-IMK, circa 2005.
2. **partners, CARROUSO.** *CARROUSO – SYSTEM SPECIFICATION AND FUNCTIONAL ARCHITECTURE.* s.l. : Yannick Mahieux - France Telecom R&D, 2001.
3. **Gardner, William G.** *3-D Audio Using Loudspeakers.* Massachusetts : Massachusetts Institute of Technology, 1997.
4. **Sima, Sylvia.** *HRTF Measurements and Filter Design for a Headphone-Based 3D-Audio System.* Hamburg : s.n., 2008.
5. **Rayleigh, Lord.** *On our perception of the direction or a source of sound.* s.l. : Taylor & Francis, Ltd. on behalf of the Royal Musical Association, 1876.
6. *Binaural Auditory Beats Affect Vigilance Performance and Mood.* **LANE, JAMES D., et al.** 2, s.l. : Elsevier Science Inc., 1998, *Physiology & Behavior*, Vol. 63.
7. *BINAURAL HRTF BASED SPATIALISATION: NEW APPROACHES AND IMPLEMENTATION.* **Carty, Brian and Lazzarini, Victor.** Como : Sound and Digital Music Technology Group, National University of Ireland, Maynooth, Co. Kildare, Ireland, 2009.
8. **Catania, Pedro.** *Biofísica de la Percepción - Sistema Auditivo.* s.l. : Universidad Nacional de Cuyo Facultad de Odontología.
9. **Horvath, Pavel.** *Animation and Sound - Review of Literature. Is Action Louder than Sound.* [Online] Horvath, Pavel, 2011. [Cited:] <http://www.pavelhorvath.net/sec1-4animation%20sound.html>.
10. 10.2 The relationship between loudness and intensity. *Open Learn - Lab Space.* [Online] Lab Space. [Cited: 10 March 2011.] <http://labspace.open.ac.uk/mod/resource/view.php?id=415687>.
11. **Wolfe, Joe.** *Helmholtz Resonance.* *The University New South Wales.* [Online] The University New South Wales, 2005. [Cited: 14 March 2011.] <http://www.phys.unsw.edu.au/jw/Helmholtz.html>.
12. **Tonnesen, Cindy and Steinmetz, Joe.** *3D Sound Synthesis . Human Interface Technology Laboratory.* [Online] 1993. [Cited: 24 January 2011.] <http://www.hitl.washington.edu/scivw/EVE/l.B.1.3DSoundSynthesis.html>.
13. **Cheng, Corey I. and Wakefield, Gregory H.** *Introduction to Head-Related Transfer Functions (HRTFs): Representation of HRTFs in Time, Frequency, and Space.* [Paper] Ann Arbor : University of Michigan, 2001.
14. **STERN, R. M., WANG, DeL. and BROWN, G.** *Binaural Sound Localization.* [book auth.] DeL. WANG and G. BROWN. *Computational Auditory Scene Analysis.* New York : Wiley/IEEE Press. , 2006.
15. **Bilbao, Stefan and Kemp, Jonathan.** *Musical Applications of Fourier Analysis & Signal Processing. Musical Applications of Fourier Analysis & Signal Processing's lecture notes.* Edinburgh : University of Edinburgh, 2010.

16. **University, Cambridge.** Convolution and Deconvolution Using the DFT. *NUMERICAL RECIPES IN C: THE ART OF SCIENTIFIC COMPUTING (ISBN 0-521-43108-5)*. 1992.
17. —. Correlation and Autocorrelation Using the FFT. *NUMERICAL RECIPES IN C: THE ART OF SCIENTIFIC COMPUTING (ISBN 0-521-43108-5)*. 1992.
18. **IRCAM.** Spatialisateur. *Forumnet - Le site des utilisateurs des logiciels de l'IRCAM*. [Online] IRCAM, 2007. [Cited: 10 March 2011.]

7 TABLE OF FIGURES

Figure 1 - Human Ear (9)	5
Figure 2- Equal-Loudness Contours	6
Figure 3- Head Related Coordinate System (11).....	6
Figure 4 - Modelled head with parameters to measure time difference (12).....	7
Figure 5 - Waveforms of low and high frequencies, and shadowing effect (4).....	8
Figure 6 – Configuration and Schematic Diagram..	12
Figure 7 – ILD for Left ear at 300 Hz.....	14
Figure 8 – ILD for Left ear at 10 kHz.....	14
Figure 9 – Correlation.	15
Figure 10 –HRIRs.	16
Figure 11 - ITD.	17
Figure 12 – HRTF of Left ear.....	18
Figure 13 – HRTF of Right.....	18
Figure 14 - Correlation and ITDs.	19
Figure 15 – ILDs.....	20

8 TABLE OF EQUATIONS

$\tau = ac \ 2\sin\theta$	Equation 1.....	8
$\tau = ac (\theta + \sin\theta)$	Equation 2.....	8
$H = Y \div X$	Equation 5.....	9
$H_z = H_{min}(z)H_{ap}(z)$	Equation 6.....	9
$X_k = n = 0N - 1xne - 2\pi jknN, \ k = 0,1, \dots, N - 1$	Equation 3	10
$g * h = G(f)H(f)$	Equation 4.....	10
$Corr(g, h)_j = GkHk *$	Equation 7.....	11

A APPENDICES

INTERAURAL TIME DIFFERENCES DATA

10,000 Hz	ITD	Angle	1000 Hz	ITD	Angle	300 Hz	ITD	Angle
	4.53515E-05	0		2.26757E-05	0		4.54E-05	0
	2.26757E-05	45		0.000113379	45		0.000748	45
	2.26757E-05	90		0.00015873	90		0.000703	90
	4.53515E-05	135		6.80272E-05	135		0.000748	135
	4.53515E-05	180		4.53515E-05	180		6.8E-05	180
	4.53515E-05	225		2.26757E-05	225		0.000794	225
	2.26757E-05	270		0.00015873	270		0.000703	270
	2.26757E-05	315		9.07029E-05	315		0.000748	315
200 Hz	ITD	Angle	2000 Hz	ITD	Angle			
	9.07029E-05	0		2.26757E-05	0			
	0.000612245	45		2.26757E-05	45			
	0.000861678	90		0.000136054	90			
	0.000634921	135		2.26757E-05	135			
	4.53515E-05	180		2.26757E-05	180			
	0.000612245	225		2.26757E-05	225			
	0.000839002	270		0.000113379	270			
	0.000657596	315		6.80272E-05	315			

INTERAURAL LEVEL DIFFERENCES DATA

200Hz	Angle	ILD	300Hz	Angle	ILD	10000Hz	Angle	ILD
	0	0.632636		0	0.47896		0	6.041979
	45	3.050599		45	2.173508		45	7.951086
	90	3.946203		90	3.340973		90	22.70789
	135	3.056786		135	3.308838		135	4.900861
	180	0.459186		180	0.66452		180	3.603786
	225	1.955769		225	1.874659		225	5.909579
	270	2.837271		270	1.847586		270	9.919714
	315	1.696232		315	1.193867		315	18.24245
1000Hz	Angle	ILD	2000Hz	Angle	ILD			
	0	1.210783		0	0.131181			
	45	12.55549		45	10.2413			
	90	7.392008		90	9.24731			
	135	8.258311		135	6.281867			
	180	0.179674		180	1.178272			
	225	8.104448		225	2.887285			
	270	5.180668		270	7.854965			
	315	9.99726		315	10.41546			