SJÄLVSTÄNDIGA ARBETEN I MATEMATIK

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET

On the Mathematics of Spatial Hearing

av

Viktor Wase

2014 - No 12

MATEMATISKA INSTITUTIONEN, STOCKHOLMS UNIVERSITET, 106 91 STOCKHOLM

On the Mathematics of Spatial Hearing

Viktor Wase

Självständigt arbete i matematik 15 högskolepoäng, Grundnivå

Handledare: Yishao Zhou

2014

STOCKHOLM'S UNIVERSITY

On the Mathematics of Spatial Hearing

Viktor Wase

Supervisor: *Yishao Zhou*

Abstract: *This paper studies the ability of spatial hearing and the synthesis of 3D audio. Some of the most well researched subjects concerning 3D audio is the shape of the position dependent frequency response of the ears (the so called Head Related Transfer Function) and the difference of a sound's arrival time between the ears (the so called Inter aural Time Difference). Using anthropological data and basic geometry a new ITD model is constructed. It is then compared somewhat favourably to the classical model of Woodworth.*
The effect of the shape of the outer ear (pinnae) on the elevation of HRTF is investigated as well. The results were inconclusive, but encouraging.
Finally a couple of different interpolations schemes were tried and investigated on the measurements from the '94 KEMAR database.

1 INTRODUCTION TO SOUND MODELING IN 3D AND HEAD RELATED TRANSFER FUNCTIONS

Most people have the ability to make out where a sound came from. This spatial hearing has three different factors[4]:
a) Difference in the arrival time of the sound.
b) Difference in magnitude in the sound.
c) Difference in the frequency content of the sound.
The difference in arrival time has been named inter-aural time differences (ITD). This is mostly owing to the simple fact that if the origin of the sound is closer to the right ear for example, then the sound wave requires a little while longer in order to reach the left ear. The mathematics behind the ITD is quite simple and will be discussed at a later point.

The mathematics behind the change in frequency is far from simple. While it is not impossible to simulate using a 3D-model of a head with torso and shoulders[1][2], the much more common way of approaching this problem is using empirical measured data. Most often using a realistic human sized doll (dummy).
This change in frequency content comes from the blockage and reflections owing to the shape of the pinnae (the outer ear), but also to a smaller degree the reflections from the shoulders and the torso[4]. Already here the observant reader can spot something that might cause trouble down the line: peoples upper bodies have vastly different shapes. This is especially true for the pinnae[8]. These differences makes it hard to know exactly how the frequency content should be treated. One might suspect that in order to work really well the procedure would have to be custom fitted to each listener.
Luckily experience has shown that it works reasonably well even without custom fitting.

1.1 KEMAR

One of the most common set of measurement come from Bill Gardner and Keith Martin[3]. They used a dummy called KEMAR (*Knowles Electronics Manikin for Acoustic Research*).
It had a proper upper half of a body with one exception: in each of its ears there was a microphone. KEMAR had two different sized ears to make up for the individual differences in ear size.
The procedure was simple: an impulse sound was played from a speaker and the impulse response (IR) in was recorded in KEMAR's ears. This kind of impulse response is often referred to as the Head Related Impulse Response (HRIR).
The speaker stayed 1.4 m from the doll and systematically varied the elevation and azimuth (horizontal angle compared to the head) and kept recording the IR. Using this spherical grid of measured points one can then use different interpolation schemes in order to approximate the points between the grid points.
In this article the KEMAR database has been used unless otherwise stated.

1.2 Assumptions in the Recording of IR

There are a lot of assumptions and approximations in the creation of these databases. We have, of course, already mentioned the difference in pinnae shape.
The KEMAR recordings tried to make up for this by using one big and one small ear-piece. While this might be a good approach it was based on a assumption of right-left symmetry. This assumption has since then been proven to be false; the right and left ear do not change the frequency content in exactly the same way[5]. The difference is however rather small, and can be neglected without making a huge impact on the result.
When one speaks about the recordings of the HRIR one has to make a distinction between the so called near-field and far-field. The line is usually drawn at roughly 1 m; all sounds closer than this are in the near-field and all sounds further away are in the far-field. Why this distinction? Because in the far-field the HRIR are close to independent of distance[6]. That means that one can measure the HRIR at a fixed range and factor in the distance at a later point as long as the distance is more than 1 m. This is what the KEMAR database has done,

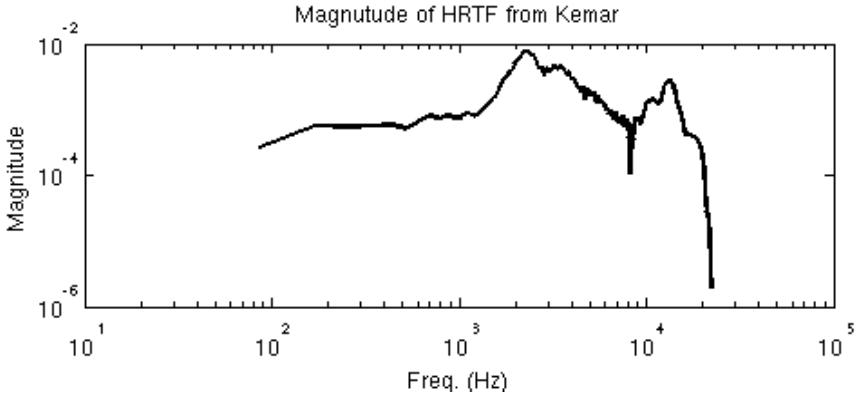

Figure 1.1: The magnitude of the measured Head-Related Transfer Function of Kemar when azimuth = 0 and elevation = 0.

and this is what I will do as well.

1.3 Head Related Transfer Function

So far we have only discussed the HRIR, but one of the main focus points of this article will be on the interpolation of the transfer function associated with the HRIR. The so called Head Related Transfer Function (HRTF). The HRTF is the Fourier transform of the HRIR.
The HRTF is often modeled as a function $H(\theta, \phi, f)$, where θ is the angle along the horizontal plane (azimuth), ϕ is the angle along the vertical plane (elevation) and f is frequency. One can by the mentioned procedure measure the HRIR for a fixed ϕ and θ, and then find the HRTF for these ϕ and θ by Fourier transfer.
The HRTF has a very specific appearance; it is rather flat in the lower frequencies (often 1kHz and below) while the higher range of the spectrum usually consists of a series of peaks and notches. See fig. 1.1. for an example.

2 Inter-Aural Time Differences

When the sound is closer to either ear there will be a delay between the ears. This is the major cue for spatial hearing.
In 1972 Woodworth proposed the following model for calculating the delay

$$ITD = \frac{a}{c}(\sin\theta + \theta)$$

where a is the radius of the head and c is the speed of sound[9]. While this is formula is attractive in its simplicity it has been shown to be inadequate for synthesis on the whole sphere. It also approximates r as infinity in order to make the model distance-independent,

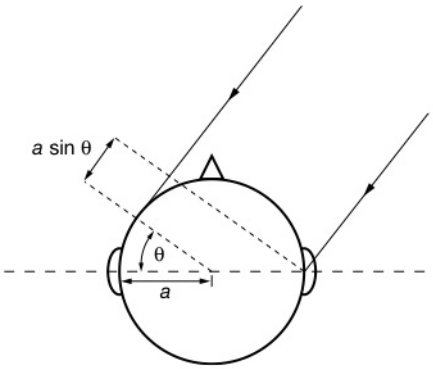

Figure 2.1: Motivation of Woodworth's model of ITD.

a trait which is desired in spatial hearing models since it usually reduces the complexity of the model. However nowadays we have an abundance of computer power at our hands, and it might be time to make accuracy the top priority when creating these models.
The way Woodworth motivated his model can be seen in fig. 2.1. (note that Woodworth use an other definition of the angle θ). Assuming that the distance is infinity gives that the lines that the sound waves propagate along are parallel. Woodworth also assumed that when the sound wave aiming for the further-most ear reached the head it smoothly changed its course and started to follow the arc of the circle (the head). While this is certainly not true for all frequencies it is an adequate model of what is happening to the higher frequencies. And since the higher frequencies are what we are mostly using for spatial hearing, it will do for now.
This model led Savioja et al. to develop the idea further[10]. They proposed the model

$$ITD = \frac{a}{c}(\sin\theta + \theta)\cos\phi$$

and found that it was a good fit to their empirical data. It has later been found to be an adequate model even in full sphere synthesis.

2.1 Distance dependent ITD model

I will now propose a slightly altered model in an effort to make it a little bit more accurate when it comes to range.
Assume that the head of the listener is a circle, and call the distance from the middle of the circle to the emission point of the sound r. Call the distance the sound has to travel in order to reach the closest ear d_c. Furthermore call the shortest possible distance from the emission point to the furthermost ear d_f.
It is clear that $d_f = l_1 + l_2$, where l_1 is the length of the circle's tangent that intersects the emission point and l_2 is the remaining arc length. See figure 2.2. for clarification.
If we start with the case $a \leq r\cos(\theta)$ (when the x-coordinate of the emission point has a larger

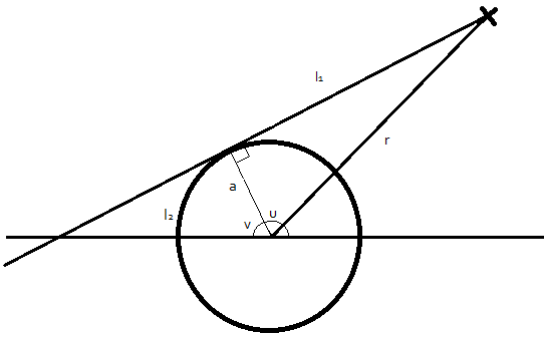

Figure 2.2: The sound's emission point is the cross and finds its closest path to the ear furthest away.

absolute value than a) using Pythagora's theorem gives

$$d_c^2 = (r\sin\theta)^2 + (r\cos\theta - a)^2 = r^2(\sin\theta^2 + \cos\theta^2) - 2ra\cos\theta + a^2$$

$$d_c = \sqrt{r^2 - 2ra\cos\theta + a^2}$$

Using the same line of reasoning in the triangle created by the lines r, l_1 and the radius of the circle gives

$$r^2 = l_1^2 + a^2 \implies l_1 = \sqrt{r^2 - a^2}$$

The arclength l_2 is av where v is the angle indicated in fig. 2.3. From the same figure one can see that the angle u can be defined as

$$u = \arccos\left(\frac{a}{r}\right)$$

This gives

$$\pi = v + u + \theta \implies v = \pi - \arccos\left(\frac{a}{r}\right) - \theta$$

With this in mind we can find the shortest distance to the furthermost ear

$$d_f = l_1 + l_2 = \sqrt{r^2 - a^2} + a(\pi - \arccos\left(\frac{a}{r}\right) - \theta)$$

The ITD is of course $(d_f - d_c)/c$.

When $|a| > |r\cos(\theta)|$ (which only happens in Woodworth's model when $\theta = \pi/2 \implies ITD = 0$ since $r \to \infty$) the shorter path isn't simply a straight line, but instead like the longer path consists of a straight line and and an arc segment. Hence this must be treated separately.
Call the angle between the x-axis and the point where the straight line from the emission point intersects the circle w. See fig. 2.3. for clarification. There exists two right triangles in the figure. Both have a hypotenuse with length r and one side of length a. Hence the third

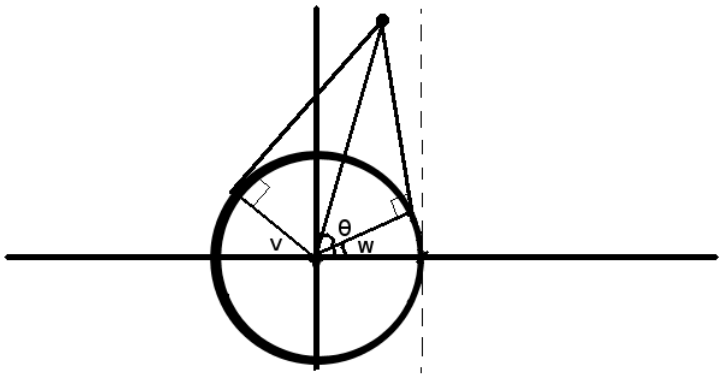

Figure 2.3: When the sound is coming from a point left of the dotted line the closest path to the closest (right) ear requires an arc.

side of the triangles are equal. This means that the ITD in this case is only dependent on the both arc lengths. That is:

$$ITD_{|a|>|r\cos(\theta)|} = \frac{d_f - d_c}{c} = \frac{a}{c}(v - w)$$

All that remains is to find w and v. Since the two before mentioned triangles are similar we can see that

$$\pi = w + 2(\theta - w) + v$$

Rearranging this gives

$$v - w = \pi - 2\theta$$

Hence

$$ITD_{|a|>|r\cos(\theta)|} = \frac{a(\pi - 2\theta)}{c}$$

2.2 Refined model of ITD

Hence my model of ITD is

$$W(\theta, r) = \begin{cases} \frac{a}{c}\left(\left(\sqrt{\left(\frac{r}{a}\right)^2 - 1}\right) + \pi - \arccos\left(\frac{a}{r}\right) - \theta - \sqrt{\left(\frac{r}{a}\right)^2 - 2\left(\frac{r}{a}\right)\cos(\theta) + 1}\right) & \text{if } |a| \leq |r\cos\theta| \\ \frac{a}{c}(\pi - 2\theta) & \text{otherwise} \end{cases}$$

One can see that Woodworth's model is a special case of this as $r \to \infty$.

$$\lim_{r\to\infty} W(\theta, r) = \lim_{r\to\infty} \frac{a}{c}\left(\left(\sqrt{\left(\frac{r}{a}\right)^2 - 1}\right) + \pi - \arccos\left(\frac{a}{r}\right) - \theta - \sqrt{\left(\frac{r}{a}\right)^2 - 2\left(\frac{r}{a}\right)\cos(\theta) + 1}\right)$$

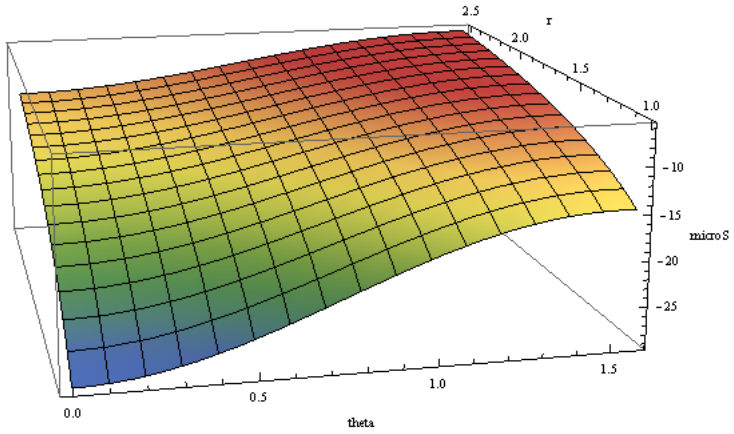

Figure 2.4: Difference between Woodworth's model and the proposed one.

$$= \lim_{r\to\infty} \frac{a}{c}\left(\left(\sqrt{\left(\frac{r}{a}\right)^2 - 1}\right) + \pi - \frac{\pi}{2} - \theta - \sqrt{\left(\frac{r}{a}\right)^2 - 2\left(\frac{r}{a}\right)\cos(\theta) + 1}\right)$$

$$= \frac{a}{c}(\cos(\theta) + \frac{\pi}{2} - \theta)$$

Since Woodworth's model used the definition $\frac{\pi}{2} - \theta$ of the angle this gives ITD = $\sin\theta - \theta$.
What about the difference between these two models? Assuming that $c = 343m/s$ and approximating $a = 0.1$ we can see that the difference in the far field ($r > 1m$) is at its peak about $25\mu S$. See fig. 2.4. for the graph of the difference. In Woodworth's formula a difference of $25\mu S$ represents an angle shift of roughly 2 degrees.

$$25 \cdot 10^{-6} = \sin\theta + \theta \implies \theta \approx 0.0425 \text{ rad} \approx 2.43^o$$

This is not much of an improvement. The sample frequency of a wave file is 44.1 kHz meaning that the sample length is about $22.6\mu S$. Since the difference between the models is usually smaller than this the difference is too small to even implement in most systems. Clearly, there isn't much use in this new model.

2.3 Cone of Confusion and Front-to-Back Confusion

Woodworth also discovered that his model led to what he chose to call the Cone of Confusion. That is a cone in which the ITD is constant. In such a cone it is often harder to determine the position from which the sound is originating.
One way to tackle this is of course to use HRTF, but even then there are problems. While the HRTF is not exactly front to back symmetric the differences are rather small. This can lead to the front-to-back confusion, which is a usual problem in 3D-audio synthesis[11].
In these cones of confusion it has been shown that the actual ITD may vary by as much as 120 μS, which is quite a lot[12]. One of the main reasons for this is the fact that the human head is not in fact a perfect circle (yes, shocker! I know.) and the ears are not placed in the

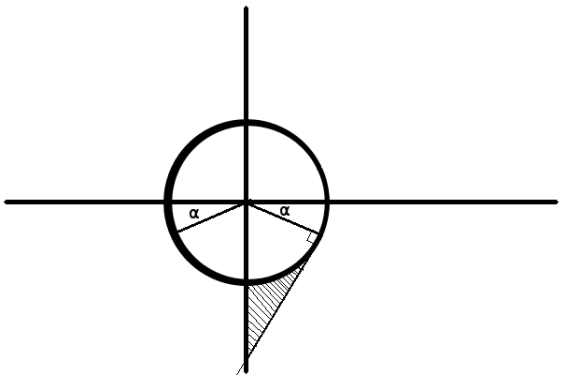

Figure 2.5: The backwards ear shift α. The nose is pointing along the y-axis.

middle of the head. Recent studies have found that when modeling the head as a 3D-ellipse moving the ears back by a just a little bit causes the synthesized ITD's to line up extremely well with empirical data. Unfortunately there is no simple analytical solution to the shortest path problem presented in the article. I will therefore propose a slightly simplified model that is quick and simple to implement while still fighting the front-to-back confusion.
Call the angle shift of the ears α. See fig 2.5. It is not as clear as before how the shortest paths from the emission point to the ears looks like.
Without loss of generality we can assume that the sound is coming from the right. In order to know how the shortest path from the point to an ear looks like one needs to know three things:

Q1: Can the closest ear be reached with only a straight line?
Q2: Is the closest path to the furthermost ear in front of the head? Q3: Can the furthermost ear be reached with only a straight line?

This leads to the classification shown in figure 2.6. We will now work through these regions one by one.

	Q1	Q2	Q3
1	Yes	No	No
2	No	No	No
3	No	No	No
4	Yes	Yes	No
5	No	Yes	No
1	Yes	No	Yes

We start in area 5. Since neither ear can be reached with only a straight line both paths must have a circle arc as well. That means that we must calculate the arclength of the path to the closest ear.
Call the angle between the x-axis and the point where the straight line from the emission point intersects the circle v. See fig. 2.7. for clarification. The shortest path from the point to the closest (right) ear is the sum of rightmost side of the triangle created by this point

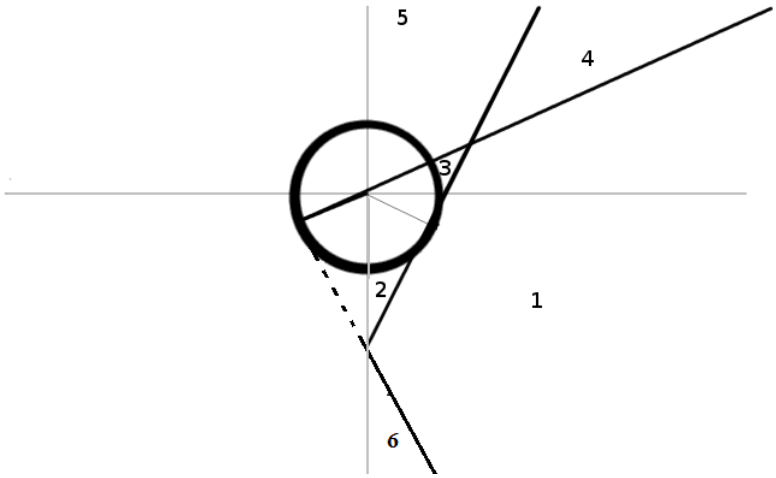

Figure 2.6: The different sections caused by the ear shift α.

of intersection, origo and the emission point and the arc length. By the same reasoning as before the side of the triangle is $\sqrt{r^2-a^2}$ and the arc length is $a(\nu+\alpha)$.
The lower left angle in the mentioned triangle is $\theta-\nu$, and we can see that $\cos(\theta-\nu)=\frac{a}{r}$. Since $0\leq\theta-\nu\leq\frac{\pi}{2}$ we get

$$\nu=\theta-\arccos\left(\frac{a}{r}\right)$$

Hence the shortest path to the closest ear d_c is

$$d_c=a\left(\sqrt{\left(\frac{r}{a}\right)^2+1}+\alpha-\arccos\left(\frac{a}{r}\right)+\theta\right)$$

When it comes to the furthermost (left) ear the calculations are exactly the same as in the previous section with the exception that the arc length is increased by $a\alpha$. Hence

$$ITD_5=\frac{d_l-d_c}{c}=\frac{a}{c}\left(\sqrt{1+\left(\frac{r}{a}\right)^2}+\pi-\arccos\left(\frac{a}{r}\right)-\theta+\alpha-\sqrt{\left(\frac{r}{a}\right)^2+1}-\alpha+\arccos\left(\frac{a}{r}\right)-\theta\right)$$

$$ITD_5=\frac{a}{c}(\pi-2\theta)$$

We continue with region 4. If the whole coordinate system is shifted by α counter clock-wise then this is almost the same as before when there was no ear shift. The only difference is that the arc to the furthermost ear is $2a\alpha$ longer. (One α for the coordinate shift and one for the shift of the other ear). Hence

$$ITD_4=\frac{a}{c}\left(\left(\sqrt{\left(\frac{r}{a}\right)^2-1}\right)+\pi-\arccos\left(\frac{a}{r}\right)-\theta+\alpha-\sqrt{\left(\frac{r}{a}\right)^2-2\left(\frac{r}{a}\right)\cos(\theta+\alpha)+1}\right)$$

Region 3 doesn't need a formula since even the most far away point is in the near-field for any sound value of α.
In the triangle in figure 2.8 we can see that using the law of sines we get

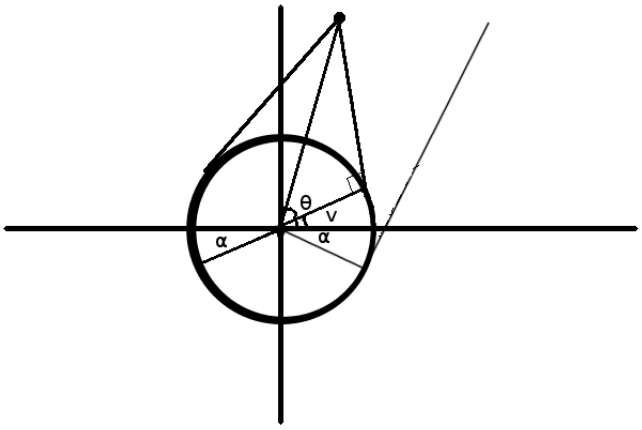

Figure 2.7: The point is the point of emission and v is the angle where the sound first hits the head.

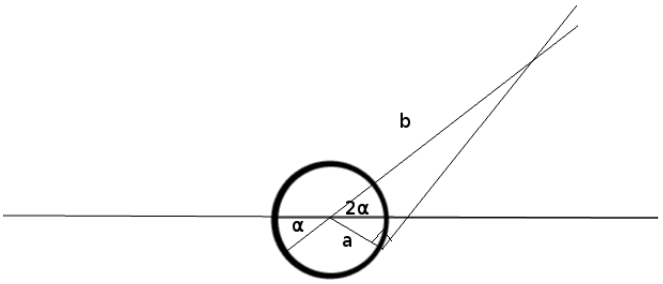

Figure 2.8: An angular shift α simplifies the calculations is section 4.

$$\frac{\sin\left(\frac{\pi}{2}\right)}{b} = \frac{\sin\left(\frac{\pi}{2} - 2\alpha\right)}{a} \implies b = \frac{a}{\cos(2\alpha)}$$

and since b must be around 1.4 m to be in the far field and a usually is around 0.1 which gives $\alpha \approx 0.75$. And this is way too much. So we ignore region 3 and move on to region 2.
Fortunately for us region 2 is similar to before when $|a| > |r\cos(\theta)|$. All that differs is a α counter clock-wise coordinate shift, and the sign of θ. This gives

$$ITD_2 = \frac{a(2(\alpha - \theta) - \pi)}{c}$$

And finally: region 1. The distance to the closest ear is easily calculated by converting to Cartesian coordinates and using Pythagoras' theorem. Using the notation from figure 9.5 we can see that

$$\pi = \alpha + v + w + \alpha$$

And with Pythagoras' once again the length straight line part of the path to the furthermost ear is $\sqrt{r^2 - a^2}$. In order to find w we can calculate this length again, but this time we use the difference in the Cartesian coordinates. Hence

$$r^2 - a^2 = (a\sin(-(w+\alpha)) - r\sin\alpha)^2 + (a\cos(-(w+\alpha)) - r\cos\theta)^2$$

$$r^2 - a^2 = a^2 - 2ar\cos(\alpha + \theta + w) + r^2$$

$$\frac{a}{r} = \cos(\alpha + \theta + w)$$

$$w = \arccos\left(\frac{a}{r}\right) - \alpha - \theta$$

Combining all these parts gives

$$ITD_1 = \frac{\sqrt{r^2 - a^2} + av - \sqrt{(r\sin\theta - a\sin(-\alpha))^2 + (r\cos\theta - a\cos(-\alpha))^2}}{c}$$

$$ITD_1 = \frac{a}{c}\left(\sqrt{\left(\frac{r}{a}\right)^2 - 1} + (\pi - \alpha + \theta - \arccos\left(\frac{a}{r}\right)) - \sqrt{\left(\frac{r}{a}\right)^2 - 2\frac{r}{a}\cos(\alpha + \theta) + 1}\right)$$

And finally: section 6. Since both ears can be reached with a straight line this is simple

$$ITD_6 = \frac{1}{c}(\sqrt{(a\cos(-\alpha) - r\cos(\theta))^2 + (a\sin(-\alpha) - r\sin\theta)^2} - \sqrt{(a\cos(\pi + \alpha) - r\cos(\theta))^2 + (a\sin(\pi + \alpha) - r\sin\theta)^2})$$

$$ITD_6 = \frac{a}{c}\left(\sqrt{1 - 2\frac{r}{a}\cos(\alpha + \theta) + r^2} - \sqrt{1 - 2\frac{r}{a}\cos(\alpha - \theta) + r^2}\right)$$

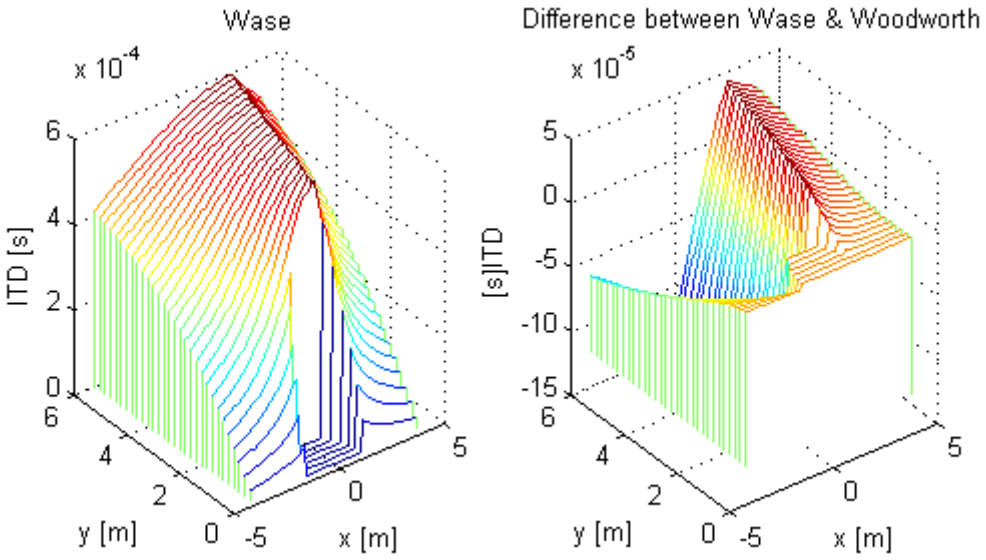

Figure 2.9: The difference between the new ITD function and Woodworth's with $\alpha = \pi/8$.

2.4 Comparison to Woodworth's model

Call this new function $ITD_W(\theta, r, a, \alpha)$. For now assume that $a = 0.08$m. It does of course wary from person to person, but a is almost always $0.05 < a < 0.1$ [16]. Also set c to $344 m/s$.
The value of α varies a lot from person to person, but according to the CIPIC database a good estimation of its mean value is roughly $\pi/5$. In the comparison below α will have therefore the values $\pi/8$, $\pi/6$ and $\pi/4$.
In the figures below the right hand side is shown, since the ITD is an odd function in x, with the near-field ($r < 1.4$) removed. The nose of the head is pointing along the y-axis.

In a normal wave file the sampling rate is 44.1 kHz. This means that the smallest ITD change that can be implemented in most systems $\approx 23\mu S$. This time the difference between the two ITD functions is greater than this for most values of the input parameters.
The difference seems to have the same general appearance for the plausible values of α. The main object of the new ITD, however, was to decrease the front-to-back confusion. This was partly successful since ITD_{Wase} is not front to back symmetric, but difference is not huge. For a lot of emission points this difference is less than $23\mu S$.

Since the KEMAR database is simply a collection of impulse responses it is possible to retrieve some ITD data from it. However this will be for a fixed distance $r = 1.4$ m. I wrote a quick script that found the ITD as a function of the azimuth θ when elevation $\phi = 0$ by finding the time where a sample was at least 15% of the amplitude of the maximum value. See fig. 2.12 for the ITD data. Then one needs to find the best parameter fit for the both ITD-models. I developed an evolutionary algorithm that found the best fit (in the least mean square sense) of the parameter a (head radius) for Woodworth's model and the parameters a and α (angular

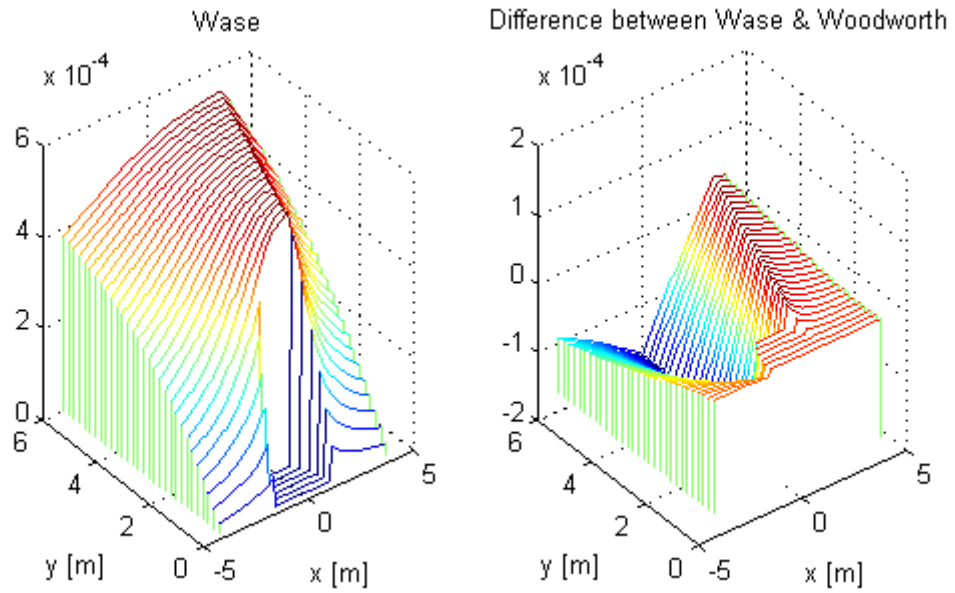

Figure 2.10: The difference between the new ITD function and Woodworth's with $\alpha = \pi/6$.

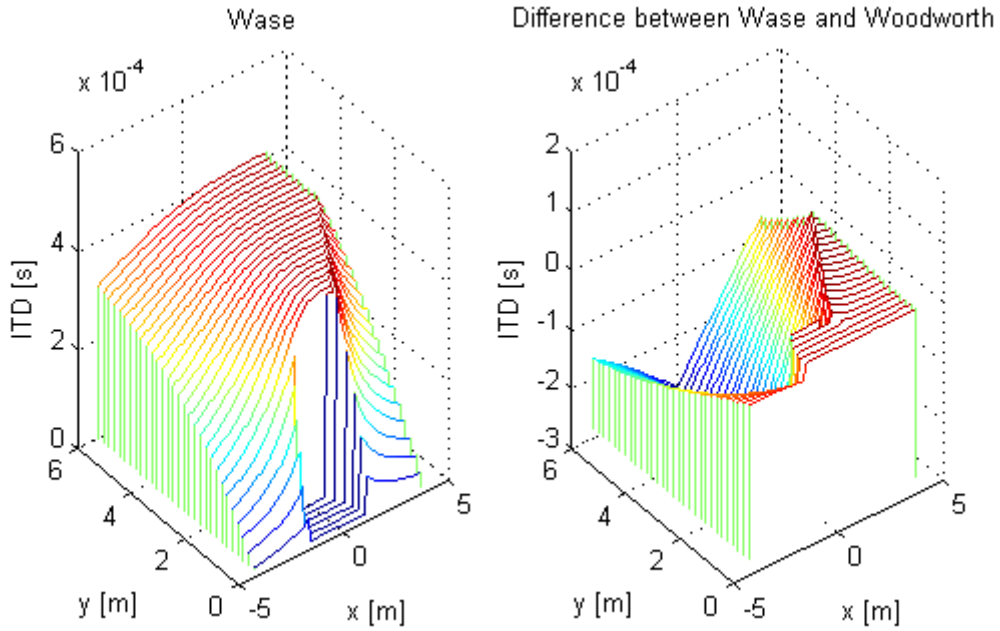

Figure 2.11: The difference between the new ITD function and Woodworth's with $\alpha = \pi/4$.

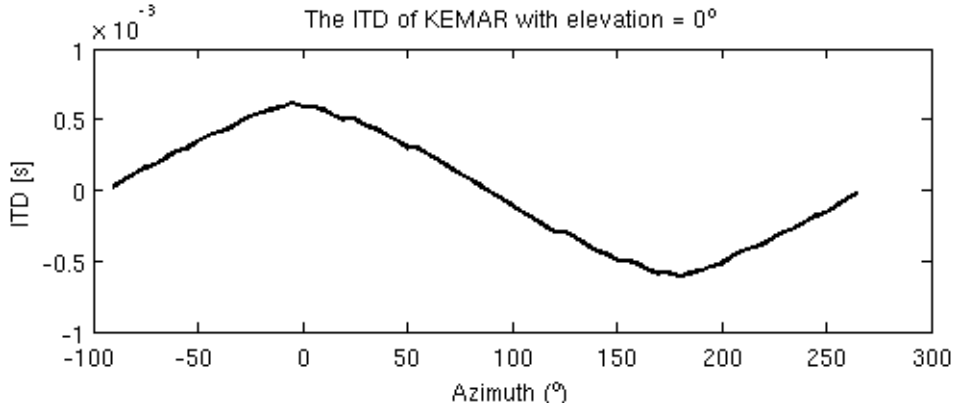

Figure 2.12: The measured ITD data from the KEMAR database.

ear shift) in my model, by randomly guessing at parameter values and updating the value if the least square error is smaller than before. I found that the value of α shrunk rather quickly through the iterations of the algorithm, thus getting closer to the model proposed in section 2.2. As stated before, that model is indistinguishable from Woodworth's in any real application owing to the sampling rate of 44.1 kHz. Hence the two models are almost equal in this setting.

However the KEMAR database only contains ITD data at r = 1.4 m, which is not enough to draw any decent conclusions from. In the 2009 paper Distance-Dependent Head-Related Transfer Functions Measured With High Spatial Resolution Using a Spark Gap T. Qu et. al created a new HRTF database using a dummy. The difference is that they have collected data for several distances.
They also used a higher sample frequency than usual, meaning that the ITD data should be slightly more accurate.

By once again using the same evolutionary algorithm to find the parameters for the best least square fit we can compare the the two models. For both models the radius of the head a converged rather rapidly to about 8 cm. In the proposed model the shift of the ears α shrunk quickly and converged to slightly below 1 degree. This is much smaller than previously thought. Even so the error shrunk a little. The average l_2-error in each of the 8 x 36 sample points for Woodworth's model was $1.30 \cdot 10^{-6}$ and my proposed model had an average error of $1.18 \cdot 10^{-6}$. The improvement was only just under 9%. Hardly much considering that a new variable r as well as a new parameter α had to be introduced.

The fact that Woodsworth's model is astoundingly accurate for its simplicity is nothing new, but the problem was of course the front-to-back confusion. In figure 2.13 the difference between the two models with the optimal parameters is shown.

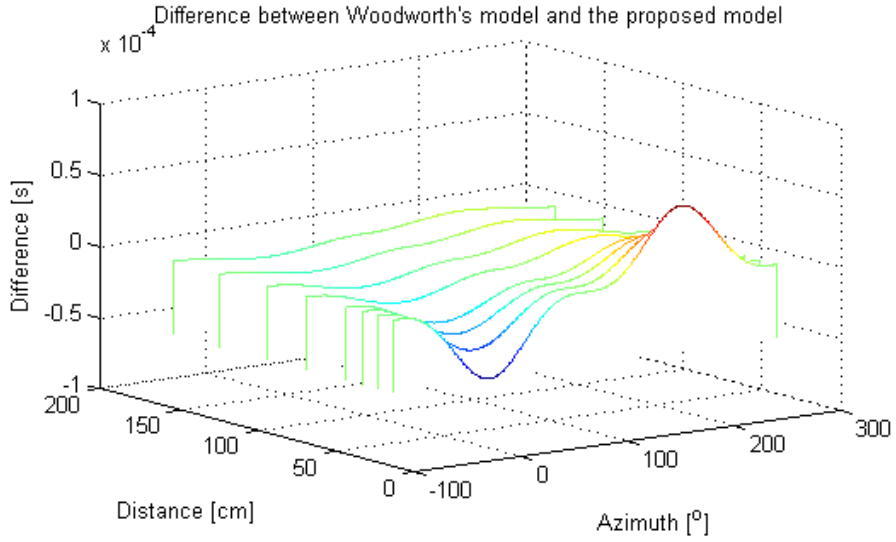

Figure 2.13: The difference between Woodworth's model and the newly proposed model.

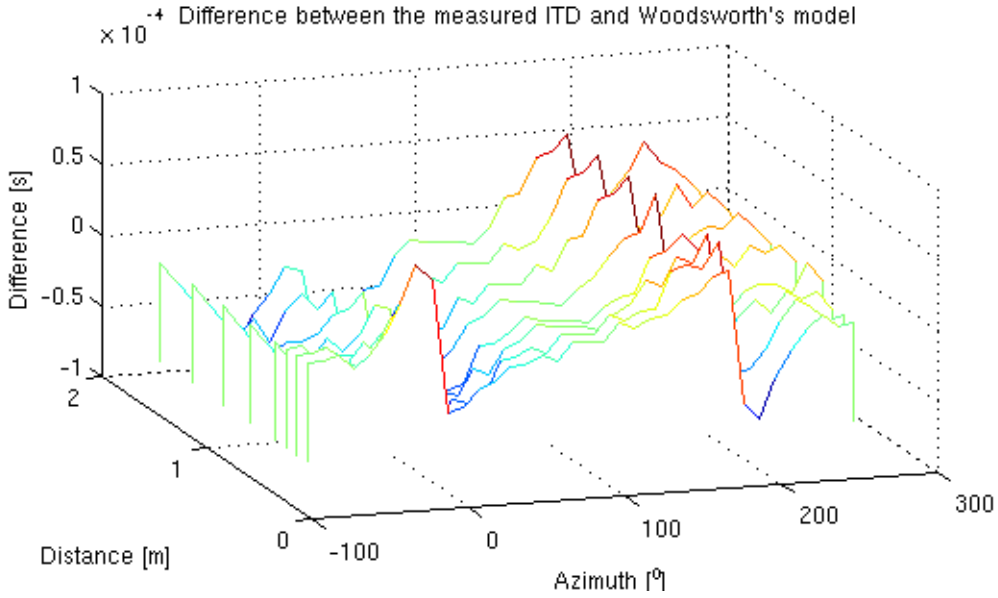

Figure 2.14: The difference between the ITD measured using a spark gap in the near field and Woodworth's model. Notice the peak at around 180^o.

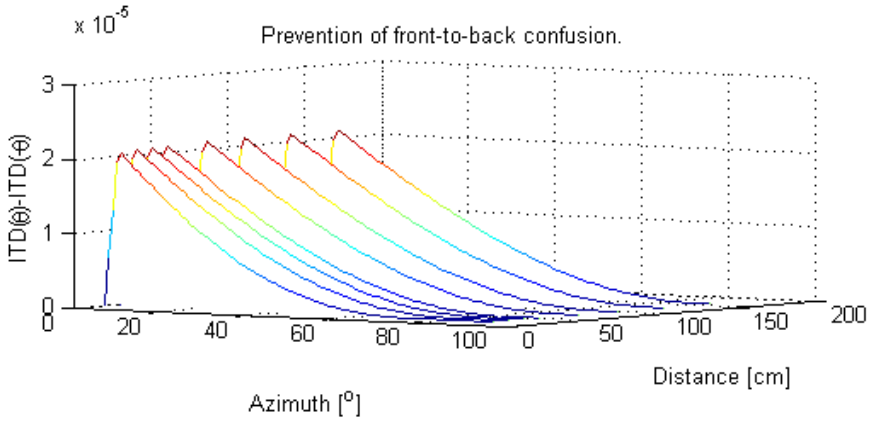

Figure 2.15: The front to back difference of the proposed model with ear shift $\alpha = 3^o$

The question is now whether or not this helps against front-to-back confusion. The largest differences are at 0 and 270 degrees, that is when the point of emission is straight to the left or right of the head. Unfortunately this is the only place where front-to-back confusion is, per definition, impossible. It is also the places where there is the least localization errors in most studies.
The ITD function is front to back symmetric iff $ITD(r,\theta) = ITD(r,-\theta)$. A way of investigating if the new model prevents this kind of symmetry is to measure $ITD(r,\theta) - ITD(r,-\theta)$. In figure 2.15 below this proof of non-symmetry is shown. As mentioned before the smallest time difference most systems can implement is around 22 μS. In order for the time difference to even get this big at its maximum α has to be $\approx 3^o$, which is several times larger than the optimal fit value.
The only possible conclusion that can be drawn from this is that the proposed model unfortunately does very little to reduce front-to-back confusion.

2.5 ITD IN ELEVATION

Up to this point we have only considered ITD as a function of two variables r and θ. But of course any proper ITD-model should depend on elevation ϕ as well.
The simplest way of achieving the 3D model would be to multiply the distance r with $\cos\phi$, since $r\cos\phi$ is the orthogonal projection of the distance vector on the plane studied in the previous sections. This is indeed what Savioja et al. did, and it has been shown to work very well with Woodsworth's planar model.
Another way is to repeat the process with the ear-shift but in 3D. This is close to what Duda, Avendano and Algazi did in their study[12]. This works great and reduces the errors in the cones of confusion significantly. The problem with that approach is that it gets complicated rather quickly. As stated before they were unable to find a closed analytical formula, but instead gave a way of solving for the desired value.
There might of course exist a region dependent formula, like the one found in the previous section. Seeing how messy the formula got in only two dimensions, however, I chose not to

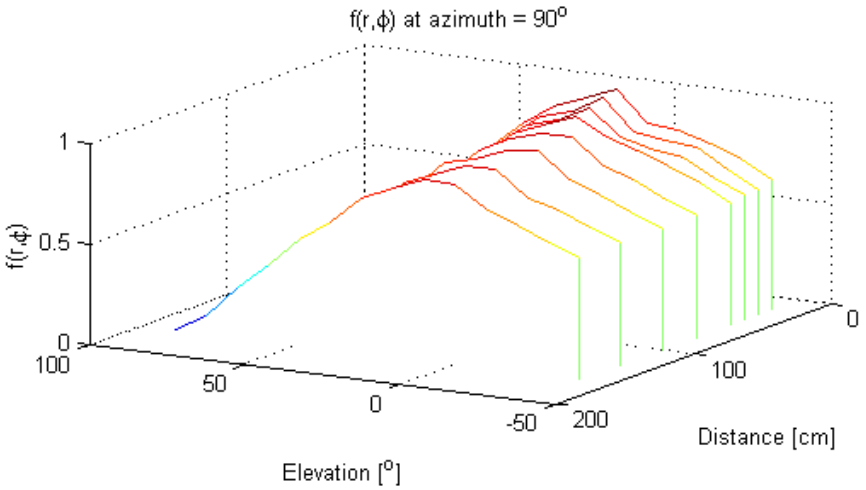

Figure 2.16: The function f that tries to separate elevation and azimuth dependence of ITD, with azimuth $\theta = 90^o$.

do this in order to preserve what little elegance and simplicity the ITD-model has.
Instead I try finding a function $f(r,\phi)$ such that it separates the variables θ and ϕ in the ITD_{3D}-function. Hence:

$$ITD_{3D}(r,\theta,\phi) = ITD_w(r,\theta)f(r,\phi)$$

The measured data will have to suffice as the ITD_{3D}-function. Hence an approximation of f might be found by inspection the behaviour of $ITD_{3D}(r,\theta,\phi)/ITD_{wase}(r,\theta)$.
In figure 2.15 we can see the result of this when the azimuth is 0^o. The graph has similar behaviour for all azimuths (except 90 and -90, when the numerical instabilities take over since the computer is asked to divide by zero), making the assumption of separation of variables seem rather plausible. And if the assumption is false it is of little importance, since it seems to lead to a good approximation.
In said graph one can see that the behaviour along the elevation axis does closely resemble a cosinus function. When θ is close to 0 and 180 there is a small error for all values of r. It is around where the elevation is 0. It has the form of a small sharp peak that is slightly higher than the value of the cosine function. It can be seen in fig. 2.15, especially when r is small. It is however small enough to be ignored without severely impacting the result.
A closer inspection of the graph shows that the f function is not independent of r. In figure 2.16 below we have the graph corresponding to $f(r,\phi)$ but at azimuth 20^o. On the right side this measured $f(r,\phi)$ is divided by $\cos\phi$ to reveal the nature of the error. Here it is clear that the function is indeed dependent of r. As r grows its influence over the function decreases, and it looks like the r-part of the function can be said to have converged when r is in the far field.
To avoid complicating things too much I will therefore consider $f(r,\phi) = f(\phi)$ to be distance independent. It should be noted that this might prevent full sphere synthesis of ITD in the near field.

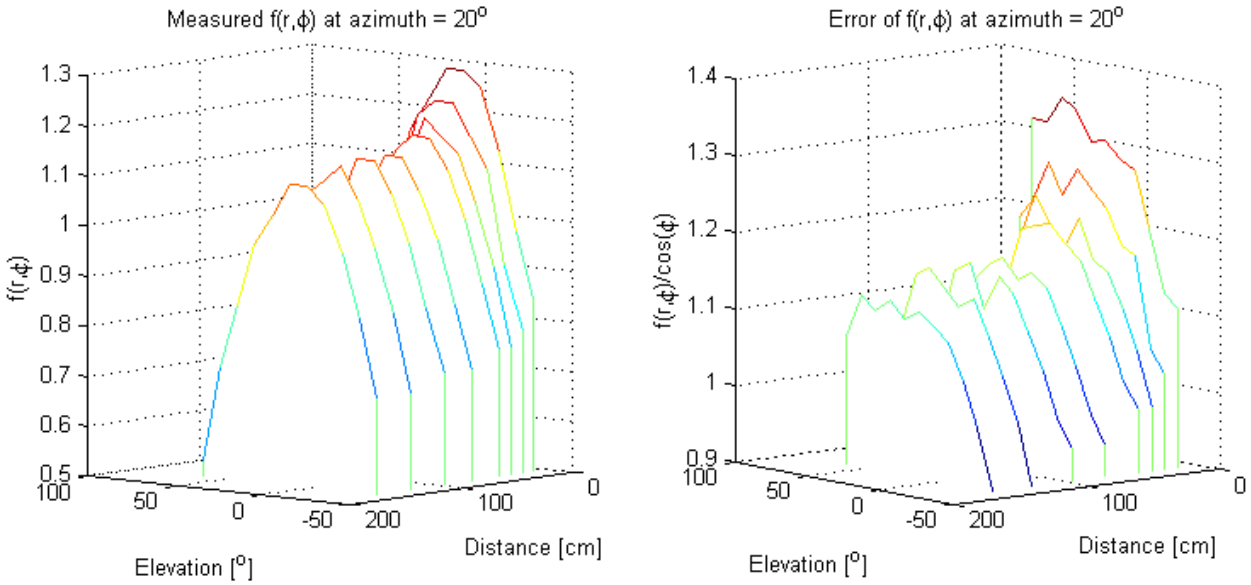

Figure 2.17: The function f (and its corresponding error) that tries to separate elevation and azimuth dependence of ITD, with azimuth $\theta = 20^o$.

3 Analysis of Recorded Head-Related Transfer Function

Some researchers have tried to model the HRTF as a smooth function that depends on a few parameters. These models are often build on the assumption that the HRTF is a combination of peak and notch filters. Simply put a peak/notch filter is a filter which has one central frequency. In the vicinity of this frequency it either amplifies the incoming data (peak) or it reduces it (notch).

These peaks and notches are (primarily) owing to the shape of the pinnae. In fig. 3.1. one can see an example of how the shape of the pinnae causes the sound wave to interfere with itself. For some frequencies this interference is destructive, other constructive; giving rise to the peaks and notches in the HRTF.

Since these creases in the pinnae vary continuously it seems natural the peaks and notches should do so too.

In a recent study Ramos et al.[13] used a second order low frequency shelving filter in combination with 12 second order peak audio filters with parameters central frequency w, log-gain G and quality factor Q, on the form

$$H_i(z) = \frac{b_0^i + b_1^i z^{-1} + b_2^i z^{-2}}{a_0^i + a_1^i z^{-1} + a_2^i z^{-2}}$$

$$a^i = [1 + \alpha_i g_i, -2\cos w_i, 1 - \alpha_i g_i]$$

$$b^i = [1 + \alpha_i / g_i, -2\cos w_i, 1 - \alpha_i / g_i]$$

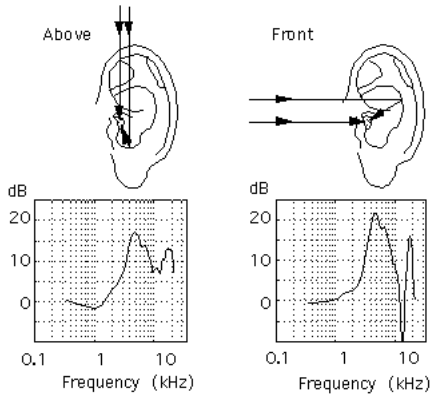

Figure 3.1: Example of how the sound waves interfere with each other at different elevations.

where $\alpha_i = \sin w_i/(2Q_i)$ and $g_i = 10^{G_i/40}$. $H(z)$ is the general form of a second order peak/notch filter, where z is the (possibly complex) frequency. w is, as mentioned before, the frequency which the filter is centered around, G is the amplitude of the filter and w is the width. This was found to be a good fit. The main points of the study was to decrease the size of the database by only storing the parameters at the measured points as well as allowing easier interpolation by reduction of parameters. However instead of simplifying the calculations of interpolation I think that a new take on this approach can make the interpolation smoother and possibly even closer to the correct HRTF.

3.1 Elevation Dependent Behaviour of the Peaks and Notches

When it comes to determining the elevation these peaks and notches are necessary since the ITD that depends on elevation tends to be almost neglectable. Since these peaks and notches vary continuously a good way to interpolate in elevation is with total respect to the peaks and notches. That is: find out how the central frequency w varies as a function of elevation ϕ for the major peaks and notches.

This must be admitted to be a long shot. There might not exist such a function. In the study On the Relation Between Pinna Reflection Patterns and Head-Related Transfer Function Features[14] they tried modelling how the peaks and notches of the elevation dependent part of the HRTF behaved based on pinna shape. They reduced the folds of the pinna to three polynomials of degree five. This study is recent and showed that there might be a new approach worth investigating.

If there does exist such a function (or a close approximation of it) then one can note that Q should be close to constant while G changes smoothly. With the reduced noise one can interpolate G using one of the usual schemes, cubic spline for example.
Call the function of the first j filters $A_j(w_{1toj}, G_{1toj}, Q_{1toj})$. By implementing an evolutionary algorithm that minimizes the error $|||A_N(w,G,Q,z)| - |HRTF(z)|||_2$ one can get a good

approximation of the parameters w, G and Q for a fixed point (θ, ϕ).
The reason for the absolute value around the functions is because they are both complex-valued functions. But since the phase-shift is irrelevant for higher frequencies, which is where the HRTF is relevant, it seemed appropriate to ignore the phase-shift.

3.2 Details of the algorithm for finding $w(\phi)$

The algorithm has two stages. In the first stage it starts with only one filter and minimizes $|||A_1(w,G,Q,z)|-|HRTF(z)|||_2$ by randomly altering w_1, Q_1 and G_1 and update them if $|||A_1(w,G,Q,z)|-|HRTF(z)|||_2$ shrinks. If the parameters have been altered 1000 times without being updated it is considered to minimized. Then another filter is added and the procedure is repeated. When there are 12 filters stage two begins. Stage two is randomly altering all 12 (w_i, Q_i, G_i) triples. If they have been altered 10000 times without improvement the program terminates. This whole procedure of stage 1 and 2 is repeated 10 times, and the fit with the smallest error is chosen. Then the exact same thing is done to all the other sampled elevations for this fixed azimuth.

3.3 Results of the Algorithm and Further Investigation

This script was computationally intense and it took a couple of days for it to run in order to get it to give reliable results. Unfortunately the results were anything but positive. It seems very unlikely that there exists even a rough approximation of the $w(\phi)$ function. There is not much similarity between any of $w(\phi)$-functions for any azimuth or elevation.

Hence I chose to give up on the $w(\phi)$ strategy and try something a little closer to the approach in the study [14]. As mentioned before a lot of the elevation dependent changes in the peaks and notches are owing to the constructive and destructive interference caused by reflections of the sound wave. Figure 3.1 shows an example of how this could work. In the study they started by finding how these folds look like by investigating a picture of the ear in question. I will instead try to find the shape of the ear from the information concerning the peaks and notches. If this shape is constant for different elevations it might be possible to use that underlying shapes to interpolate more smoothly.

3.4 Reflections in the Pinna

The fact that different people have different ear sizes and different ear shapes has proved to be a slight problem in 3D audio sythesis. If one could find some strong connection between the shape of the HRTF and the shape of the ear it could increase the quality of the personalization of the HRTF. That is why I shall try to reverse engineer the shape of the ear from the peaks and notches.

Imagine that a sound source is infinity far away. All the sound waves will act like parallel rays. One of these rays will head right for the ear canal. Call this line l_1. There will be another

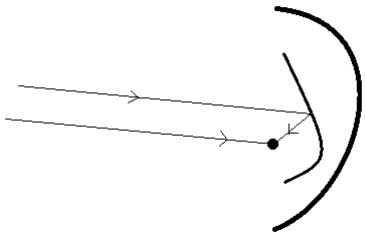

Figure 3.2: Example of how the shape of the pinna might cause destructive/constructive behaviour in the frequency response.

line l_2 which will bounce of one of the folds of the ear and then make its way to the ear canal. See figure 3.3 below for a graphical explanation.

Clearly l_2 needs to travel longer than l_1. Call this difference Δl. It is somewhat likely that the ray which is most strongly reflected is the one whose angle of incidence is equal to its angle of reflection. That is, we assume that the ear has a rather smooth surface. Call this incidence angle v, and its complementary angle $w = \frac{\pi}{2} - v$. Call the function representing the first ear fold $r_1(\theta)$ (this function is given in polar coordinates owing to the concave shape of the fold). This point of reflection is where the line l_2 intersects r_1. In order to find the either the angle v or w one can find the slope of the ear r_1'. And using the difference between the slope of the line which is orthogonal to r_1' and the slope of l_1 we can get the angle v. But the angle of the slope of l_1 is per definition ϕ

$$v = -\tan(\arctan(r') - \phi)$$

If one defines the coordinate system with the origo in the ear canal then the hypotenuse of the right angle triangle in the figure 3.3 is of length $r_1(\theta_0)$, where θ_0 is the point of intersection. Call this length d. The total difference Δl is the sum of d and the side through which the sound enters the ear. Basic trigonometry then gives

$$\Delta l = d(1 + \cos(2w))$$

But when does this give rise to peaks and notches? The first thing to notice is that the reflection coefficient is most likely negative[14]. This means that there will be a peak at frequency f_n when the difference is

$$\Delta l = \frac{c(n+1)}{f_n}$$

where c is the speed of sound and n any natural number. In the same way there will be a notch when

$$\Delta l = \frac{c(2n+1)}{2f_n}$$

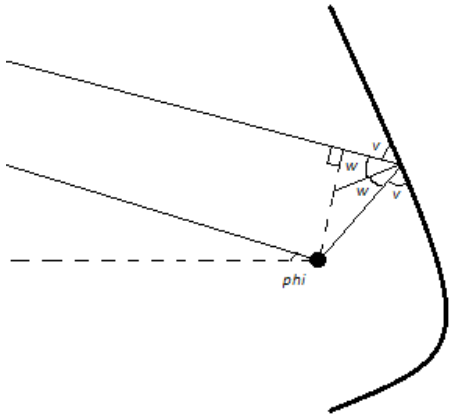

Figure 3.3: If the difference in length of the two lines is a multiple of $\lambda/2$ constructive or destructive interference will occur.

3.5 Evolutionary Algorithm to Find the Shape of the Pinna

Using the above equations one can of course find the frequencies of the peaks and notches for a given ear shape formula, such as r_1. Assume that the ear can be reduced to three such functions $r_1(\theta)$, $r_2(\theta)$ and $r_3(\theta)$[14] . I will now construct an algorithm that finds the shapes of these three functions in order to make them fit to the already measured peaks and notches. Let $r_i(\theta)$ start as a a constant function, thus giving a circle in polar coordinates. In order to evaluate the fitness of these functions all the peaks and notches corresponding to the functions are found by iterating through the polar coordinate θ. Frequencies above 20 kHz are of course ignored. Then the algorithm iterates through the newly found peaks and notches. For each of those a value e_i is found. It is the distance to the closest measured frequency multiplied by the corresponding log gain. That is

$$e_i = w_i|G_i|$$

All of these e_i are then added together to form e.
This is not enough to minimize e since this might be frequencies w that are not covered by any peaks and notches caused by any $r_i(\theta)$. Hence the algorithm iterates through all w_i and adds the distance to the closest f_n to e.
Add a random smooth function to each of the $r_i(\theta)$ functions. If the inverted fitness e of the new functions is less than for the previous functions the functions the functions are updated. This procedure repeats for 10^5 iterations.

3.6 Result

It must be noted that while the pinna is responsible for much of the elevation dependent change in the frequency response it is not the only cause of peaks and notches in the HRTF. In figures 3.4 and 3.5 below the shape of the ear with azimuth 90 and elevation 30^o and 60^o

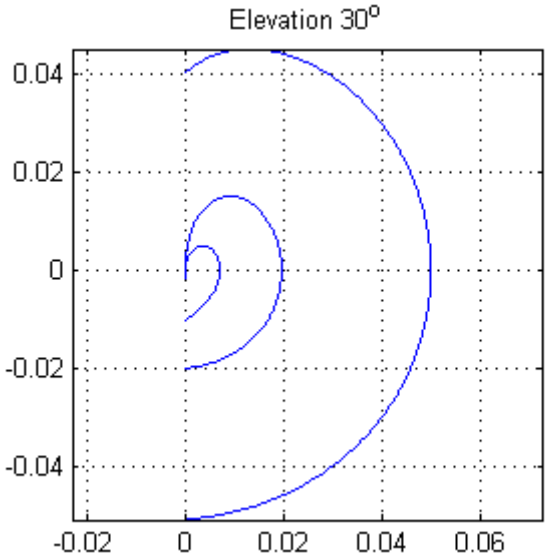

Figure 3.4: Simulation of the folds of the ear with elevation = 30 and azimuth = 90

are shown. While these two are just examples most elevations (with azimuth 90^o) have this general appearance. They are not identical, but they do have the typical shape of an ear.
Perhaps a better way of doing it is to iterate through all possible elevations for a specific azimuth and find the best possible $r_i(\theta)$ functions. In this case the ear might get a slightly more realistic shape since usually a human ear doesn't change shape when observed from different angles.
In figures 3.6 and 3.7 we see two examples of this with azimuths 90^o and 120^o. Unsurprisingly the one with azimuth 90^0 share many characteristics with fig. 3.4 and 3.5. However fig. 3.6 has a smeared appearance. While this is not perfect it does show that something is indeed working, since a flat surface becomes increasingly stretched when viewed from an angle. Unfortunately it loses the shape in this smearing process. This probably means that some of the underlying assumptions of the model were a little extreme. In figure 3.8 below the azimuth is 0^o. At this point the model breaks down almost completely. This is not all that surprising since azimuth was assumed to be 90 in the design of the model and algorithm.

It seems that this model was a little too simple to fit reality. In spite of this simple visual inspection of the ears with the sound emission point straight forward showed promising results. The shapes did indeed look like ears, and more so; they looked rather similar no matter which values the elevation had.
This type of reverse engineering of the ear could with a little luck (and more than a little hard work) lead to custom HRTF by simply taking a photograph of the ear. The individual differences in the ear's frequency response is one of the biggest problems in today's spatial audio synthesis.

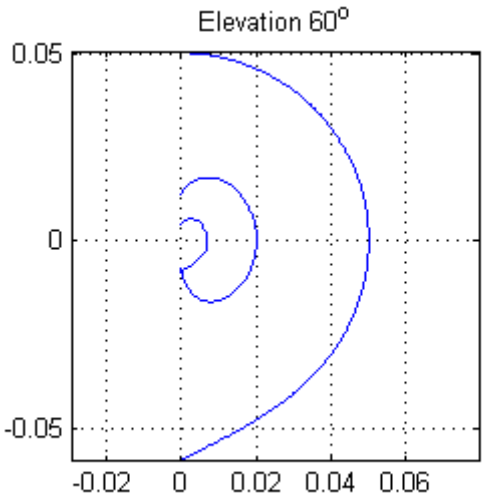

Figure 3.5: Simulation of the folds of the ear with elevation = 60 and azimuth = 90

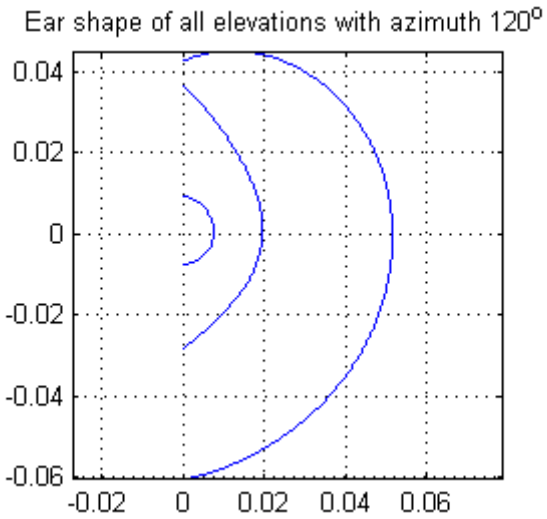

Figure 3.6: Simulation of the folds of the ear all elevations and azimuth = 120

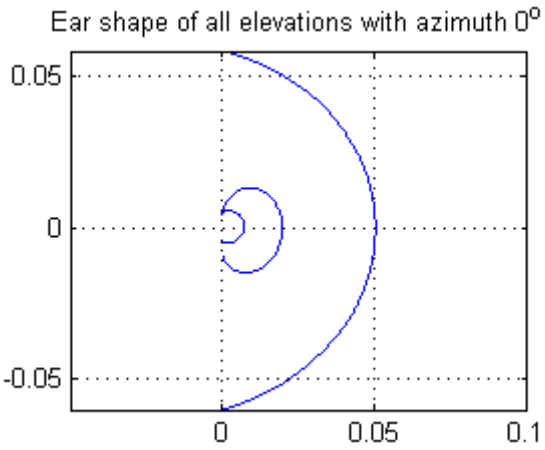

Figure 3.7: Simulation of the folds of the ear all elevations and azimuth = 90

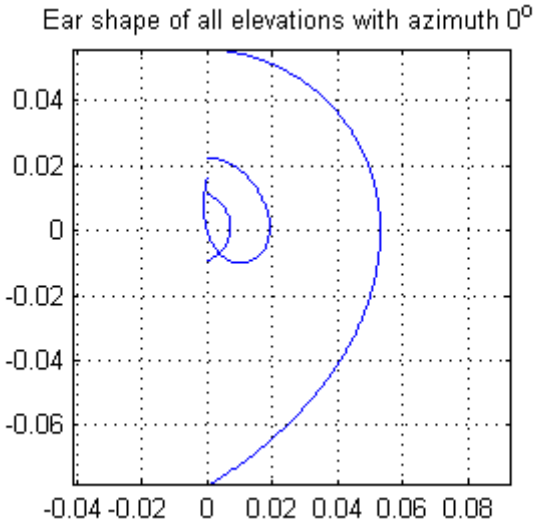

Figure 3.8: Simulation of the folds of the ear all elevations and azimuth = 0

4 INTERPOLATION

When it comes to interpolation of the HRTF along the elevation ϕ the choice of interpolation method was motivated by the fact that most of the changes were due to the shape of the outer ear (pinnae). Interpolation along the azimuth θ is slightly trickier since there isn't only one factor responsible for most of the change. Therefore one has to rely on the more common types of interpolation. In this section four different types of interpolation will be compared to each other.

4.1 LINEAR INTERPOLATION

Linear interpolation is probably the most commonly used in these sorts of interpolations since it is by far the simplest and fastest method of interpolation.

4.2 CUBIC SPLINE

After trying linear interpolation a intuitive approach would be to change the line to a higher order polynomial. However if this is done in a naive way by taking an n-order polynomial and fitting it through $n+1$ points the interpolated function will oscillate widely between the points when n is high. This is called Runge's phenomenon. In order to avoid this we continue to interpolate with piece-wise polynomials, but add some smoothness requirement.
The Spline interpolation makes the resulting curve as smooth as possible. That is, given a function $f(x)$ it minimizes the curvature k

$$k = \frac{f''(x)}{(1+f'(x)^2)^{3/2}}$$

The Spline interpolation can of course be of any order n, but the most common is the cubic and it is also the one that shall be used here. That means that the smoothness requirement is that $f_i''(\theta)$ is continuous.

4.3 PERIODIC INTERPOLATION

One thing that can be said to be true about the complex system that is HRTF is that it is 2π-periodic. This makes the choice of a periodic interpolation scheme seem intuitive. If the sampled points are exact and the function is band limited to the so called Nyquist frequency then this one might even get the exact interpolation. This is most likely not the case, and its repercussions will be discussed later.
A Fourier polynomial p of order n has the general form

$$p(x) = a_o + \sum_{k=1}^{n} a_k \cos(kx) + \sum_{k=1}^{n} b_k \sin(kx)$$

For a fixed value on r and ϕ call $HRTF(\theta,\phi,r,f) = H_f(\theta)$. Since $H_f(\theta) = H_f(\theta \pm 2n\pi), \forall \theta, f$ and $\forall n \in \mathbb{N}$ we know that there exists a (of possibly infinite order) Fourier polynomial p such

that the error between p and H_f in the L^2-norm is zero. This error is denoted $||p - H_f||_2$ with definition

$$||p - H_f||_2 = \int_{-\pi}^{\pi} (p(\theta) - H_f(\theta))^2 d\theta$$

It is worth noting that this doesn't necessary imply point-wise convergence. Take for example the function

$$sq(x) = \begin{cases} 1 & \text{if } 1 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

owing to the jump at $x = 1$ and $x = 2$ $p(2) \neq H_f(2)$ even if the order of p goes to infinity. This is called Gibb's phenomenon. But if the true HRTF is continuous, as it should be, this shouldn't be much of a problem.

The Nyquist frequency is half of the sampling frequency. The Nyquist−Shannon sampling theorem says that if the highest frequency in the sampled material is B then with a sample rate of $2B$ the original signal can be uniquely determined.
In practice the sample rate should always be a little more than $2B$ to avoid errors. This is why the sample rate of most music formats is 44.1 kHz, since the upper limit of human hearing is 20 kHz. This means that if the curve is band-limited under this frequency the interpolation will be exact. If it is not, however, there is a chance of oscillation between the sampled points. The naive way of finding the coefficients a_0 to a_n and b_1 and b_n would perhaps be to take the Fourier Transform of the curve of the HRTF at a specific elevation and frequency. Then using the properties

$$F(\cos(2\pi f_0 t)) = \frac{\delta(f - f_0) + \delta(f + f_0)}{2}$$

$$F(\sin(2\pi f_0 t)) = \frac{\delta(f - f_0) + \delta(f + f_0)}{2i}$$

the coefficients can be found easily. This way is not numerically stable and small errors tend to grow.
A much better way of doing it would be using Lagrange polynomials. An ordinary Lagrange polynomial is a linear combination of k l_j-functions on the form

$$l_j(x) = \Pi_{0 \leq m \leq k} \frac{x - x_m}{x_j - x_m} \text{ s.t. } j \neq m$$

where x_m are the sample points. The coefficients are simply the corresponding y-values, giving the interpolating function

$$p(x) = \sum_{j=0}^{k} y_j l_j(x)$$

But in order to adjust this to trigonometric polynomials the $l_j(x)$ functions are replaced by trigonometric functions $t_j(x)$ such that

$$t_j(x) = \Pi_{0 \leq m \leq k} \frac{\sin(0.5(x - x_m)}{\sin(0.5(x_k - x_m))}$$

with $k \neq m$ if there are an odd number of sample points. If there are an even number of sample points then

$$t_j(x) = \frac{\sin(0.5(x-\alpha_k)}{\sin(0.5(x_k-\alpha_k))}\Pi_{0\leq m\leq k}\frac{\sin(0.5(x-x_m)}{\sin(0.5(x_k-x_m))}$$

where

$$\alpha_k = \sum_m x_m, m \neq k$$

4.4 Comparison

In order to compare the interpolation techniques half of the sample points were discarded for the time being. Using the different interpolations approximate values to these discarded points one can get an estimation of the error.

Before we actually start comparing errors the norm must first be considered. Since human hearing is logarithmic both in pitch and loudness a logarithmic error norm should be used. We shall simply take the logarithm of the vectors before we start applying the usual error norms. While this might give us a better appreciation of the perceived error it unfortunately loses the properties that makes it a norm. For example for $p(v)$ to be a norm $p(av) = |a|p(v)$, where a is a scalar. But if

$$p(v) = n(\log(v))$$

where $n(v)$ actually is a norm, then

$$p(av) = n(\log(av)) = n(\log a + \log v) \neq |a|n(\log v)$$

So we should look into some usual norms first just to make sure that the loss of scalability doesn't affect the result too much.

In the table below one can see the number of sample points from the KEMAR dummy on each elevation. These numbers where chosen to make sure that the distance between the points were roughly equal ($\approx 0.1m$) since the circles were smaller for more extreme elevations. While this makes sense it makes it harder to interpolate when the number of sample points get too low. This is especially true for the periodic interpolation owing to the Nyquist−Shannon sampling theorem. It seems reasonable that the highest frequency at the lower elevations shouldn't be all that different from the highest frequency at the more extreme elevations since the it is the same factors in play. But according to the sampling theorem the highest frequency that we can handle is proportional to the number of sample points.

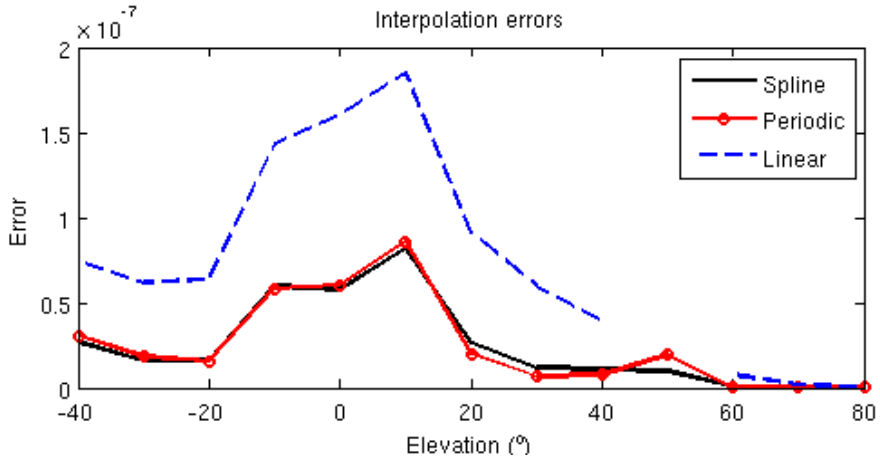

Figure 4.1: The interpolation errors measured in the l^2-norm.

Elevation	Num. of sample points
-40	56
-30	60
-20	72
-10	72
0	72
10	72
20	72
30	60
40	56
50	45
60	36
70	24
80	12
90	1

In figure 4.1. below we can see the error measured in the usual l^2-norm without compensation for logarithmic hearing. All the interpolation errors seem to have to same overall behaviour; with larger values for elevations close to 0 and smaller otherwise. It should be noted that this is not simply owing to the increased number of sample points at those elevations, since the error is divided by the number of sample points. Call the number of points at a specific elevation n, then the error is

$$e_{l^2}(x) = \frac{\sqrt{\sum_{i=1}^{n} x_i^2}}{n}$$

Already at this point it is easy to discard the linear interpolation unless speed is of paramount importance since its error is often twice as big as the other interpolation schemes. When it comes to the comparison between periodic and spline it is a closer race: more careful analysis

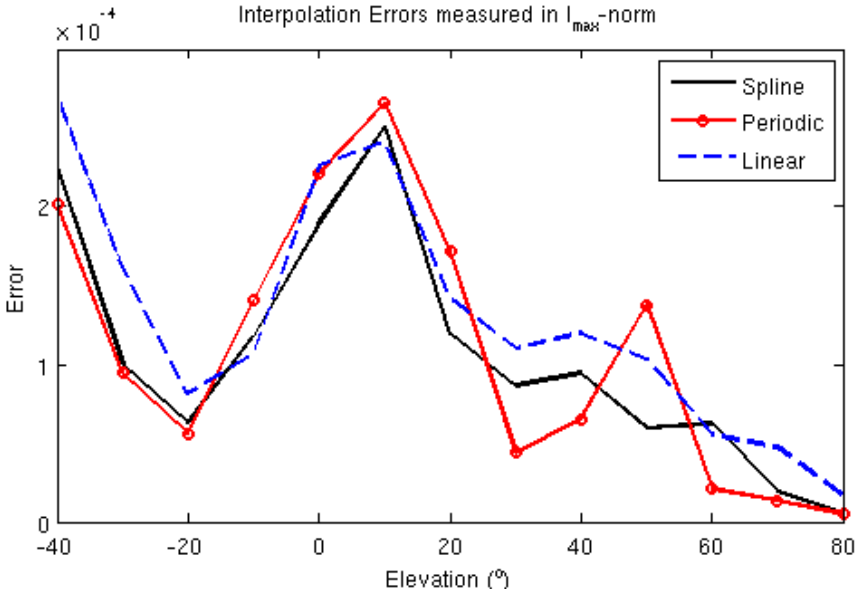

Figure 4.2: The interpolation errors measured in the l^∞-norm.

is needed.
Let's look at the l^∞-norm. It is defined as the l^p-norm as $p \to \infty$, where the l^p-norm is

$$\left(\sum_{i=1}^{n} |x_i|^p \right)^{\frac{1}{p}}$$

Put simpler it is the element in the vector x with the largest magnitude. This is important because it discovers weak spots. Convergence to zero in the continuous L^2-norm does not imply point-wise convergence, but convergence in the continuous L^∞-norm does. While we are not measuring a continuous function it is always worth t have in mind that even a small l^2-norm between two functions might have a rather big difference and one point, if the number of points is big.
The max-error is shown in figure 4.2. below. Here something interesting appears: all interpolation schemes give roughly the same result. The linear error is perhaps slightly bigger than the other ones, but not nearly as much as in l^2-norm. At elevation 50 we can see that the periodic interpolation has a clear spike (this spike is visible in the l^2-norm as well, but it is much smaller there). This is owing to the odd number of sample points at that elevation.

And finally the measurements using a logarithmic scale. Hopefully the logarithmic error should look somewhat similar to the other errors, since it is most like how the human brain perceives the sounds. In figure 4.3 below we can see that is closely resembles the l^2-error. Unfortunately this provides no further information about which interpolation scheme should be used: Fourier polynomials or cubic Spline.

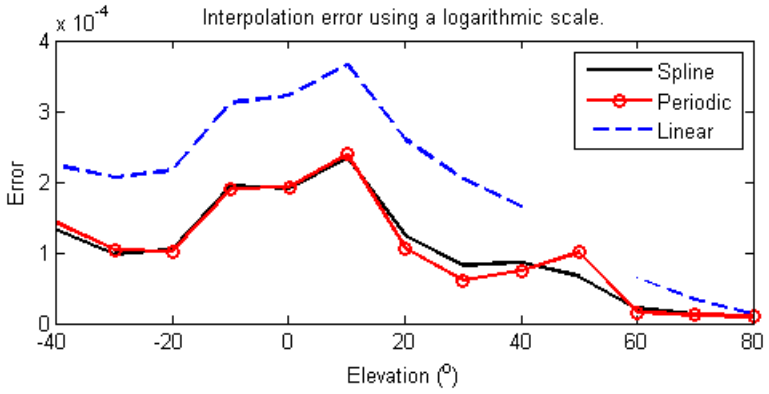

Figure 4.3: The interpolation errors measured in the l^2-norm using a logarithmic scale.

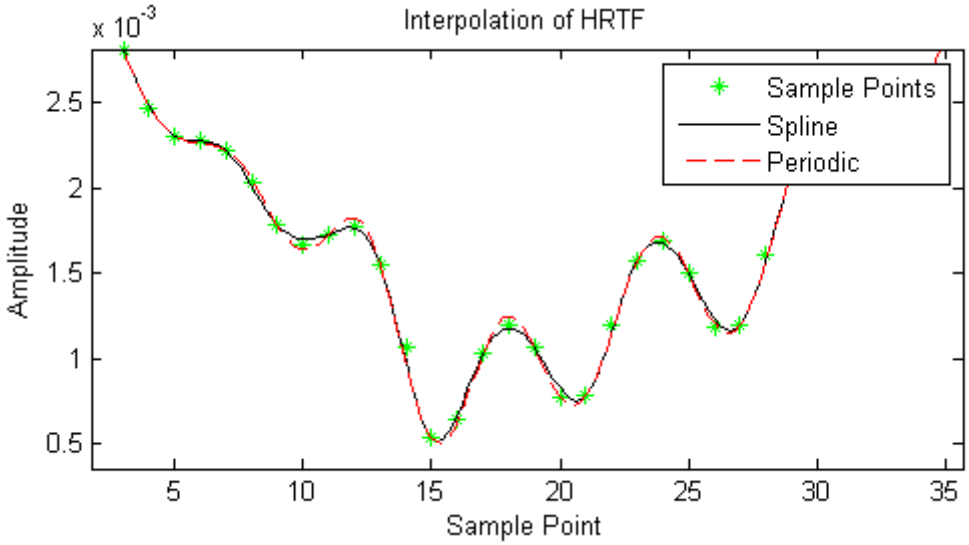

Figure 4.4: Example of the different interpolation schemes with ele = 0^o and azi = 195^o.

4.5 Noise Reduction

The problem with the periodic interpolation is that when the function isn't band limited to the Nyquist frequency the interpolant might begin to oscillate widely between sampling points. Since the sample rate was halved when every other point was removed it might actually be the case that the true sample rate might be enough. Unfortunately we would need more sample points in order to test this.
Another thing that might cause these oscillations is noise, since the noise is most likely not band limited to the Nyquist frequency either. For example will one small perturbation on one of the samples in one of the sample points look like a delta function in the time domain, thus completely making a mess of the higher frequencies in the frequency domain thus causing oscillations. In order to make this effect less prominent one could either remove the noise or ignore it.
In an effort to increase the signal-to-noise quotient the signal is convoluted with the rectangular function

$$rect(t) = \begin{cases} 1 & \text{if } |t| < 1 \\ 0 & \text{otherwise} \end{cases}$$

This is also called a moving average. Hence the new function is

$$\int_{-\infty}^{\infty} H_f(\phi - \tau) \frac{rect(s\tau)}{s} d\tau$$

where s is the scaling factor of the moving average. But since our H-function is discrete a much simpler way of doing this computation is to simply take the average of the point in question and s points to the left and s points to the right. However the points with a distance less than s from the endpoints pose a small problem, since the vector is of finite length. Hence these points will retain their original value.
See figure 4.4 for an example where $s = 1$.

In an attempt to ignore the noise one more interpolation scheme is tried. Instead of first smoothing the data we try finding a function that is smoother than the data itself. The simplest way of doing this is perhaps loosening the restriction that forces the function to pass exactly through the measured data points. Let's instead find a best fit in a least square sense. Once again piece-wise cubic spline will be used but each piece-wise polynomial will pass through n points, where n can be any integer of my choosing. See figure 4.5 for an example.

The downside with both of these approaches is that it is much harder to measure the error. This is because the way that the error was measured in the previous section assumed that the sampled data points were correct. In this case we are assuming the existence of some noise, which means that if the interpolated function doesn't pass through the removed data points it could mean that the noise was successfully removed. It could, however, also mean that the interpolation scheme failed miserably.

Perhaps the best way of testing this is by actually trying it out by letting people estimate direction and range based on hearing cues. But this is outside the scope of this paper. Instead I

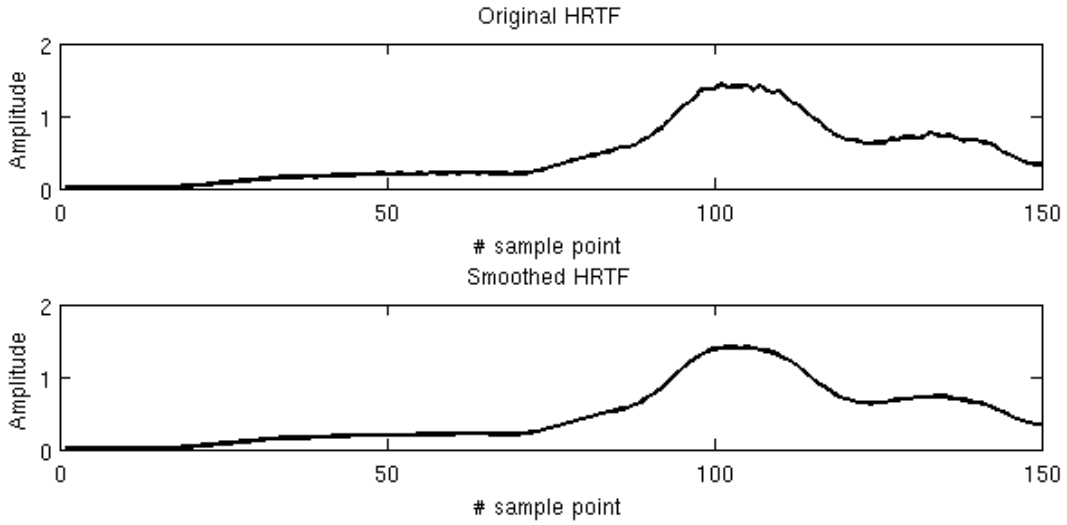

Figure 4.5: Comparison of the original function and the smoothed function with azi=90, ele=0 and s=1

have chosen to measure the curvature of the function. While this probably won't give definite proof either way it is a good first hint of the quality of the interpolation.

Since the spline actually minimizes the curvature it will be used as the base line for the measurement of the error of the trigonometric smoothing scheme. Since the trigonometric interpolation are most likely to capture the underlying mechanisms of the HRTF it seems plausible that if the trigonometric interpolation, for some smoothing factor s, has a small curvature it will accurately capture the behaviour of the true HRTF.

The total curvature will be the numerical integral of k with numerical derivatives defined

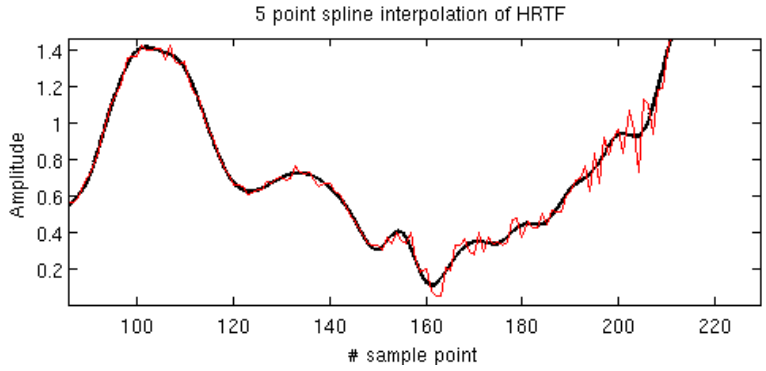

Figure 4.6: Example of 5 point spline interpolation with azi=90 and ele=0

as

$$f'(x) \approx \frac{f(x+h) - f(x-h)}{2h}$$

$$f''(x) \approx \frac{f(x+h) - 2f(x) + f(x-h)}{h^2}$$

where h is the step-size. These were chosen since they are both simple and of second order.

When it comes to deciding the values of the scaling factor s and the number of points for the piece-wise polynomial to be interpolated through n it is clear that bigger values will give smaller curvature. That is why both are made to be as small as possible. Visual inspection (for example in fig. 4.4) indicates that $n = 5$ gives a very smooth function that stays true to the shape of the original function.

Call the curvature of the spline with $n = 5$ for k_{sp}. By numerically integrating the $k(x)$ curve function of the trigonometric interpolation using $s = 0, 1, 2$ and comparing this to k_{sp} the following is found:

s	$\int k dx / k_{sp}$
0	11,068
1	3,066
2	2,032

The interpolating function clearly oscillates less than before the smoothing process. When $s = 2$ the curvature is just twice as large as the minimized curvature k_{sp}, which leads me to believe that this interpolation scheme might be worth investigating. To further increase the value of s would probably cause the function to deviate from the HRTF too much.

5 Conclusions and Further Work

5.1 Combining the Interpolations

Unfortunately the investigation of elevation dependent interpolation didn't yield any results that are straight away usable. This means that until further work on the area has been done the choice of interpolation will have to be solely based on the work done in section 4. Fortunately this makes the combination of the elevation and azimuth much simpler. Instead of finding a subtle way of making them work together one can now simply choose an interpolation scheme and generalize it up a dimension.

5.2 Further Work

The work on the ITD in section 2 yielded results that are only slightly better than the classical Woodworth model. Except when it comes to synthesis of spatial hearing in the near field and horizontal plane. In this special case of a special case my model excels. However using the anthropomorphic data provided in the CIPIC database[16] and statistical regression one could try to find the one or two parameters that have the biggest effect on the ITD. Then using

the same approach of simple geometry one could perhaps find a better ITD model that give robust results even in the near field.
The model might also be expanded to three dimensions. This would however cause the calculations to become messy, seeing how the number of sections was rather high already in 2D with only one parameter. It might, however, make for a sightly better real time ITD calculator.

The work in section 3 seem to tap into a rich vein of future research. The model used in this paper is, as stated before, much to simple to give any more than a vague glimpse of the underlying mechanisms. Perhaps one could remove the assumption that the ear is a smooth surface and thus instead of a straight line which the sound wave is reflected along one would have a distribution of lines. The larger the angle between the original reflected wave and the one in the distribution that actually passes through the ear canal, the smaller the reflection constant (in magnitude, not with sign).

The area of interpolation schemes in HRTF is a well studied area, and except for actually trying the proposed schemes on actual people there is little I could suggest to the enrich this field of research.

6 Pseudo Code

6.1 ITD

The ITD-function returns the time difference between the right and left ear in seconds. If the point of emission is closer to the left ear then the answer will be negative.
Before showing how to implement the ITD mentioned in section 2 we must first find a way of finding out which of the six regions that the sound is coming from (see fig 2.6 for a reminder). The line that goes through origo is the simple one: if $\theta > \alpha$ then the point of emission is above this line and hence in region 4 or 5.
When it comes to the other line it is clear that its slope if $\pi/2 - \alpha$ radians. By writing the equation of the line $y = kx + m$ we get $k = \tan(\pi/2 - \alpha)$. The line must pass through the point $(x, y) = (a\cos(-\alpha), a\sin(-\alpha)) = (a\cos(\alpha), -a\sin(\alpha))$. Plugging in this we can find m:

$$-a\sin\alpha = \tan(\pi/2-\alpha)a\cos\alpha + m$$
$$-\tan\alpha = \tan(\pi/2-\alpha) + \left(\frac{m}{a\cos\alpha}\right)$$
$$m = -a\cos\alpha(\tan\alpha + \tan(\pi/2-\alpha))$$
$$m = -a\cos\alpha(\tan\alpha + 1/\tan\alpha)$$

Hence we find that the eq. of the line is

$$y = \frac{x}{\tan\alpha} - a\cos\alpha(\tan\alpha + 1/\tan\alpha)$$

And hence if we denote the Cartesian coordinates of the emission point (x_{ep}, y_{ep}), then if $\tan(\pi/2-\alpha)x_{ep} - a\cos\alpha(\tan\alpha + 1/\tan\alpha) < y_{ep}$ the point is below the line.

However when α gets close to zero this gets numerically unstable. This is because $1/\tan\alpha \to \infty$ when $\alpha \to 0^+$. The computer is then asked to subtract infinity from infinity. Not the best of ideas. Let's reformulate the equation.

$$y = \frac{x}{\tan\alpha} - a\cos\alpha(\tan\alpha + 1/\tan\alpha)$$

$$y = \frac{\cos\alpha\, x - a\cos\alpha\left(\frac{\sin^2\alpha}{\cos\alpha} + \cos\alpha\right)}{\sin\alpha}$$

$$y = \frac{\cos\alpha\, x - a\left(\sin^2\alpha + \cos^2\alpha\right)}{\sin\alpha}$$

$$y = \frac{\cos\alpha\, x - a}{\sin\alpha}$$

Not only is this more elegant; when α gets close to 0 the factor $1/\sin\alpha$ gets larger until it hits the maximum floating point value the machine can handle. At this point the right hand side is larger than y_{ep} if and only if $x - a$ is positive (or non-negative if $y_{ep} < 0$). This is the same division of regions as in the range dependent model without the ear-shift α. There the model made the distinction when $a \le r\cos\theta$, but since $r\cos\theta = x_{em}$ this is the same.

function ITD(θ,ϕ,r,a,α)
$x \leftarrow r\cos\theta$
$y \leftarrow r\sin\theta$
$c \leftarrow 344$
if $\theta > \alpha$ **then**
if $\frac{\cos\alpha x - a}{\sin\alpha} < y$ **then**
$t \leftarrow \pi - 2\theta$
else
$t \leftarrow \sqrt{(r/a)^2 - 1} + \pi - \arccos(a/r) - \theta + \alpha - \sqrt{(r/a)^2 - 2(r/a)\cos(\alpha+\theta) + 1}$
end if
else
if $\frac{\cos\alpha x - a}{\sin\alpha} < y$ **then**
$t \leftarrow \pi + 2\theta$
else
if $y < \frac{\cos(\alpha-\pi/2)x - \sqrt{r^2 - 2\frac{r}{a}\cos\alpha + a^2}}{\sin(\pi/2-\alpha)}$ **then**
$t \leftarrow \sqrt{1 - 2\frac{r}{a}\cos(\alpha+\theta) + r^2} - \sqrt{1 - 2\frac{r}{a}\cos(\alpha-\theta) + r^2}$
else
$t \leftarrow \sqrt{(r/a)^2 - 1} + \pi - \arccos(a/r) + \theta - \alpha - \sqrt{(r/a)^2 - 2(r/a)\cos(\alpha+\theta) + 1}$
end if
end if
end if
return $(a/c)\,t\cos\phi$
end function

7 References

[1] Mingsian R. Bai and Teng-Chieh Tsao, 2006, *Numerical Modeling of Head-Related Transfer Functions Using the Boundary Source Representation*, J. Vib. Acoust vol 24, p594-603

[2] Ling Tang, Zhong-Hua Fu amd Lei Xie, 2013, *Numerical Calculation of the Head-Related Transfer Functions with Chinese dummy head*,IEEE Xplore

[3] Bill Gardner and Keith Martin, 1995, *HRTF Measurements of a KEMAR Dummy-Head Microphone*,J. Acoust. Soc. Am. 97 pg. 3907-3908

[4] Kalle J. Palomaki and Hannu Tiitinen, 2005, *Spatial processing ni the human auditory cortex: The effects of 3D, ITD and ILD stimulation techniques*,Brain Res. Cogn. Brain Res. pg. 364-79

[5] Xiao-Li Zhong, Feng-chun Zhang and Bo-Sun Xie, 2013, *On the Spatial Symmetry of Head Related Transfer Functions*, Applied Acoust. vol 74, pg 856-864

[6] Jyri Huopaniemi and Klaus A. J. Riederer, 1998, *Measuring and Modeling the Effect of*

Source Distance in Head-Related Transfer Functions Proceedings of the ICA/ASA '98 Conf. Seattle, 20-26.6

[7] Takanori Nishino, Seiichiro Hosoe, Kazuya Takeda and Fumitada Itakura, 2009, *Measurement of the Head Related Transfer Function using the Spark Noise*, Audio, Speech, and Language Processing, IEEE Transactions on, vol 17 pg 1124-1132

[8] Michele Geronazzo, Simone Spagnol and Federico Avanzini, 2011, *A Head-Related Transfer Function Model for Real-Time Customized 3-D Sound Rendering*, Signal-Image Technology and Internet-Based Systems (SITIS), 2011 Seventh International Conference on, pg 174-179

[9] Woodworth et al., 1972, *Woodworth and Schlosberg's Experimental phsycology*

[10] Savioja, L. ; Huopaniemi, J. ; Lokki, T. ; Vinnen, R., 1999, *Creating Interactive Virtual Acoustic Environments*, JAES vol 47, pg 675-705

[11] Akihiro Kudo, Hiroshi Higuchi, Haruhide Hokari and Shoji Shimada, 2006, *Improved method for accurate sound localization*, Acoust. Sci. and Tech vol 27, pg 134-146

[12] Richard o. Duda, Carlos Avendano and V. R. Algazi, 1999, *An Adaptable Ellipsoidal Head Model for the Interaural Time Difference*, Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on, vol 2, pg. 965-968

[13] German Ramos and Maximo Cobos, 2013, *Parametric head-related transfer function modelling and interpolation for cost-efficient binaural sound applications*, J. Acoust. Soc. Am. 134, pg 1735-8

[14] Simone Spagnol, Michele Geronazzo, and Federico Avanzini, 2013, *On the Relation Between Pinna Reflection Patterns and Head-Related Transfer Function Features*, IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL 21, pg 508-519

[15] P. Satarzadeh, 2006, *A study of physical and circuit models of the human pinnae*

[16] V. R. Algazi, R. O. Duda, D. M. Thompson and C. Avenando, 2001, *The CIPIC HRTF Database*, Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the, 99-102