Chapter 5

Numerical models of horn loaded loudspeakers

The accurate and timely prediction of far field acoustic pressure is important in the design of horns. This chapter compares results from experiments on two representative horn loaded loudspeakers with the numerical techniques described in the last chapter.

It has been found that the source superposition technique is capable of modelling these horns accurately over a wide range of frequencies, and is fast enough to become a component of a further optimisation technique.

5.1 Introduction

The source superposition technique of Koopmann and Fahnline (1997), when combined with a number of calculation speed enhancing modifications (described and verified in Chapter 4), is an ideal candidate for calculations of the beamwidth of acoustic horns. It is able to accurately model the far-field pressure from a spherical cap mounted on the surface of a sphere. The technique is not limited to a one dimensional approximation

such as those traditionally used to model horns (Holland et al., 1991, McLean et al., 1992, Mapes-Riordan, 1993), which has been shown to be invalid above a certain limiting frequency in Chapter 3.

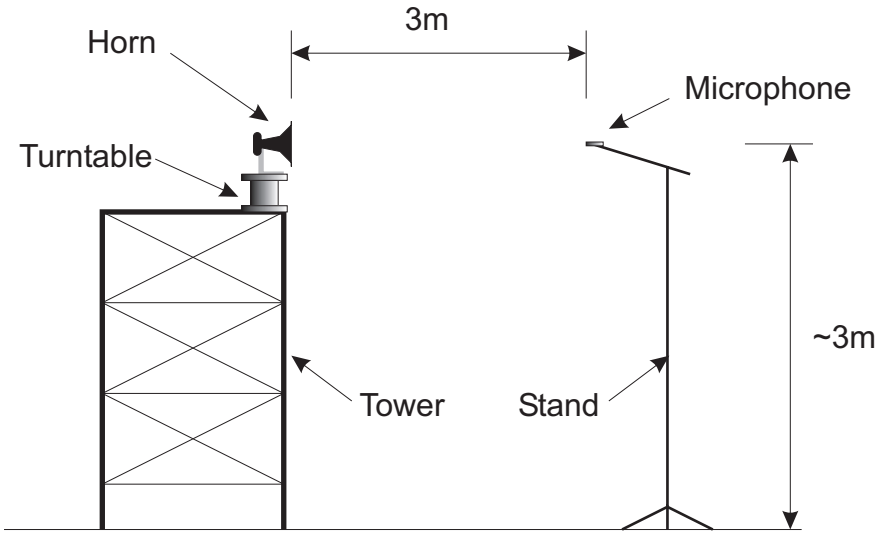
This chapter aims to validate (Babuska and Oden, 2004) the source superposition numerical model by comparison to experiment, and to alternative numerical methods, such as the direct BEM of Wu (2000). First, the experimental setup used is described and the result obtained reported. These results are then compared to those obtained by the direct BEM and the standard source superposition boundary element method over a frequency range limited by computational time of the direct BEM. Experimental results over a larger frequency range are then compared to the source superposition method with the modifications described in the previous chapter. Finally conclusions are drawn as to the utility of numerical modelling of horn loaded loudspeakers using the source superposition technique.

5.2 Experiments

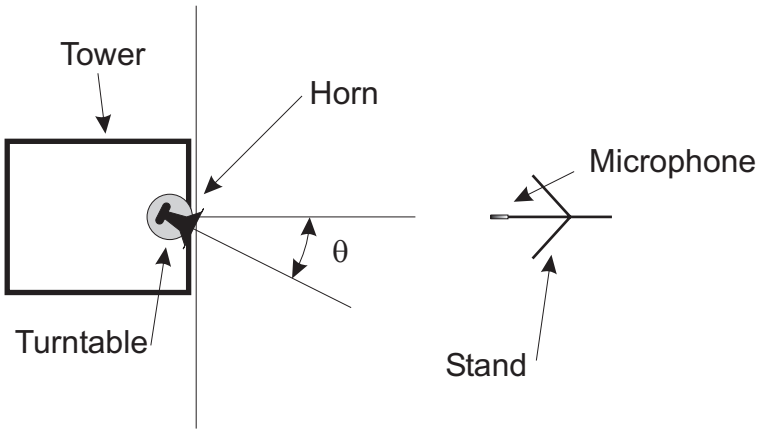
The unbaffled horns described in Section 3.2.2 were placed on an indexed rotating platform (turntable) on an elevated tower inside a large open space. The sensitivity, or pressure frequency response for 2.828 Volt rms input (1 W rms into 8Ω), of each horn was measured at a distance of 3 m from the centre of the mouth of the horn in 5° intervals ranging from on-axis (0°) to 90° off-axis. Figure 5.1 shows a diagram of both side and plan views of the setup.

At each frequency of interest, a polar plot of the magnitude of the measured acoustic pressure, normalised by the maximum pressure, was produced. Figure 5.2 shows the sound field of the two step conical horn at three different frequencies: 550 Hz, which is a low frequency for this size horn and shows a wide beam of sound; 2000 Hz, which shows a narrowing of the sound field; and 4600 Hz, which shows a beam pattern with an on

axis null, and is evidence that a velocity distribution other than that corresponding to the plane wave mode exists at the horn mouth. These experimental results give impetus for the development of accurate numerical models of horn loaded loudspeakers.



(a) Side view



(b) Plan view

Figure 5.1: Experimental setup for measuring beamwidth. The horn sits on an indexed turntable on a large tower. The sound pressure is measured at a large distance from the horn while the horn is rotated in 5° *intervals.*

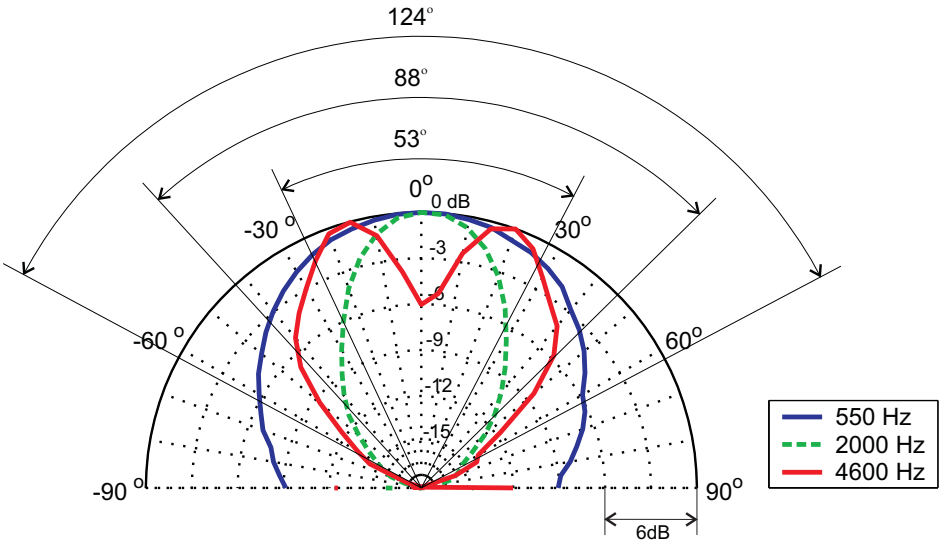


Figure 5.2: Polar plot of the magnitude of the measured pressure, normalised by the maximum pressure at that frequency, for a two step conical horn at three different frequencies.

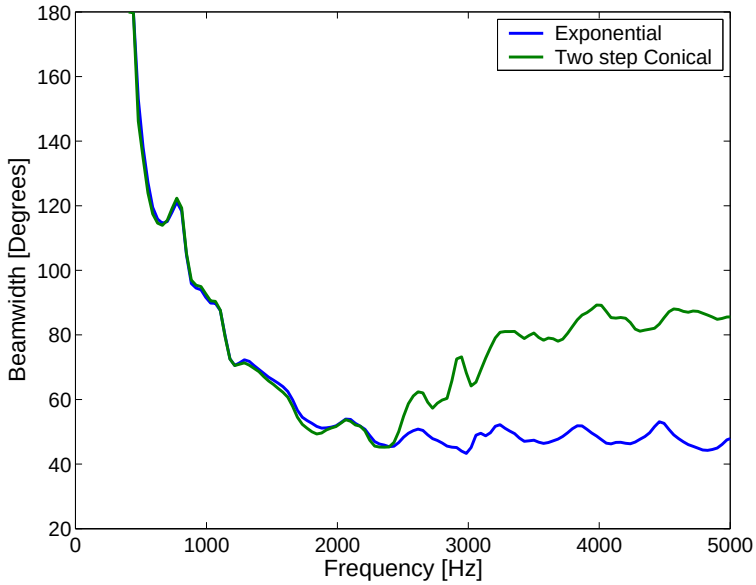


Figure 5.3: Experimental measurements of the variation of beamwidth with frequency for exponential and two step conical horns.

5.3 Comparison to standard numerical methods

Simulations of both the exponential and two step conical horns have been undertaken for both the direct BEM and the standard source superposition techniques. Figure 5.4 shows the surface mesh used to discretise the exponential horn, at a nominal 6 elements per wavelength. Figure 5.5 shows the same for the two step conical horn. Note the quarter symmetry, to reduce computation time, and the need for the horn to be placed in an artificial cylindrical volume for the direct BEM. This requirement is discussed in Section 4.2.3 in detail (see Figure 4.6). A small volume is placed over the rear of the horn throat in the source superposition mesh to stop sound radiating out the rear of the horn. A unit velocity was placed at the throat of the horn, represented by the darker areas in Figures 5.4 and 5.5.

As found in Chapter 4, the number of variables to be solved is critical for the performance of the numerical methods based on boundary element techniques. The direct BEM method discretises each element with a linear variation between each node and the number of variables is the number of nodes in the mesh. The source superposition technique places a discrete source at the centroid of each element, and the number of variables to be solved is the number of elements. Table 5.2 compares the number of variables to be solved for the direct BEM and the source superposition technique for both the exponential and two step conical meshes. Matrix inversion is an order N^3 operation (Golob and Van Loan, 1996), and a speedup of between 3 and 5 would be expected based on the reduction in mesh size alone (i.e. excluding matrix assembly time). This is a significant advantage of the source superposition technique in modelling thin structures such as horns. The number of elements used in the direct BEM method could be reduced by decreasing the size of the exterior volume. How to do this automatically is not clear, as the obvious solution of a “thin” surface that conforms to the shape of the horn introduces numerical issues (Martinez, 1991), see Section 4.2.3. Of course other numerical techniques such as the direct mixed method of Wu (1995) and the variational indirect method of Hamdi

and Ville (1986), and Vlahopoulos and Raveendra (1998) could be used. Availability of source code, along with no clear speed advantage, excluded the first choice and the need for a time consuming double integration excluded the second method.

Method	Exponential	Two step conical
Direct BEM	1216	1105
Source Superposition	849	631

Table 5.1: Comparison of matrix size produced by the direct BEM and source superposition technique.

The beamwidth of the horns was calculated for frequencies from 300 to 5000 Hz at 50 Hz intervals. The upper frequency was chosen to limit the run time required for the direct BEM method. Figure 5.6 shows a comparison with experimental results for both direct BEM and the source superposition method. The agreement between both methods and experiment is excellent. The trends exhibited by the experimental results are captured by both numerical methods. Both horns exhibit different beamwidth behaviour above a certain frequency, and this difference gives confidence that these techniques are capable of modelling the far field response from a horn of arbitrary geometry. There is a larger difference between the beamwidth calculated by the numerical techniques here than was found in the previous chapter (see, for example Figure 4.43). The reason for this difference is the different meshing strategies needed for each method when modelling thin structures such as these horns.

The standard source superposition technique was found to produce results approximately 13 to 20 times faster than the direct BEM, with the simulations run on an Intel P4 1500 MHz with 512 Mb of RAM running Windows XP.

In conclusion, this section has shown that numerical models are capable of reproducing the sound field generated by horn loaded loudspeakers from a specification of the horn geometry. The accuracy of the reproduction is adequate for design purposes within the given frequency range. Both the direct BEM and the standard source superposition

Method	Total	Time / Freq	Factor
	[seconds]	[seconds]	
Direct BEM	27096	285	13
Source Superposition	1984	21	1

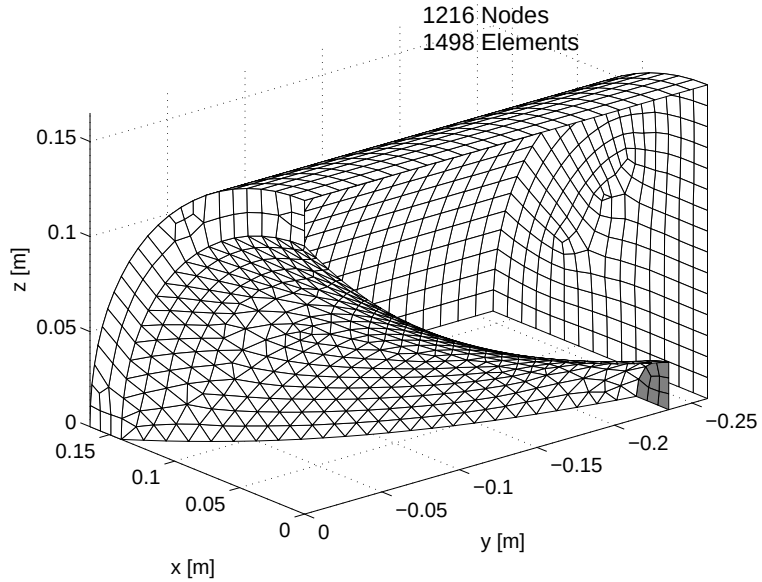
(a) Exponential

Method	Total	Time / Freq	Factor
	[seconds]	[seconds]	
Direct BEM	21137	222	20
Source Superposition	1055	11	1

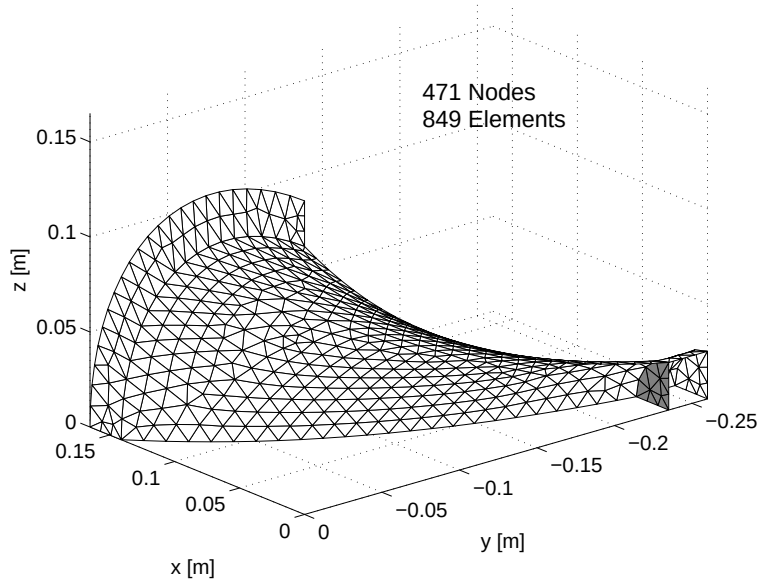
(b) Two step conical

Table 5.2: Comparison of computational time using the direct BEM and source superposition techniques, 6 *elements per wavelength at* 5000 *Hz.*

technique are capable of reproducing the experimental beamwidth, however the source superposition technique is considerably faster. A calculation time of 5 to 7 hours for 95 frequencies, such as that required for the direct BEM, is not feasible for inclusion in any sort of practical optimisation routine.

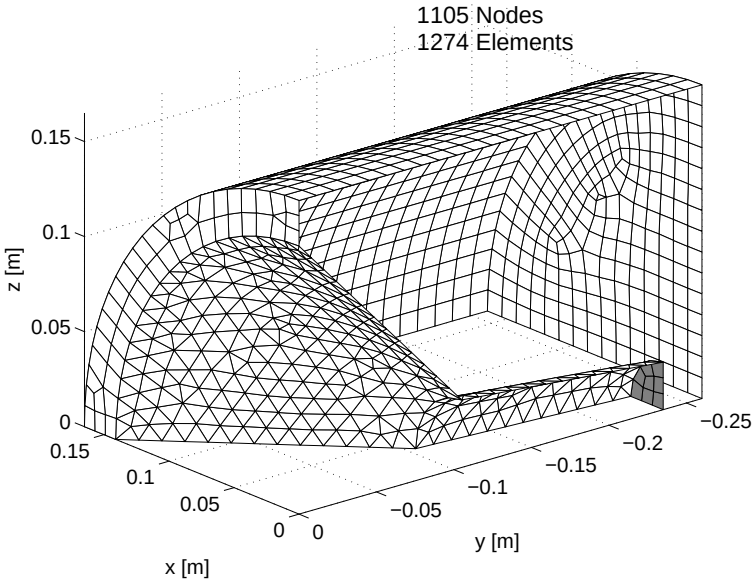


(a) Direct BEM

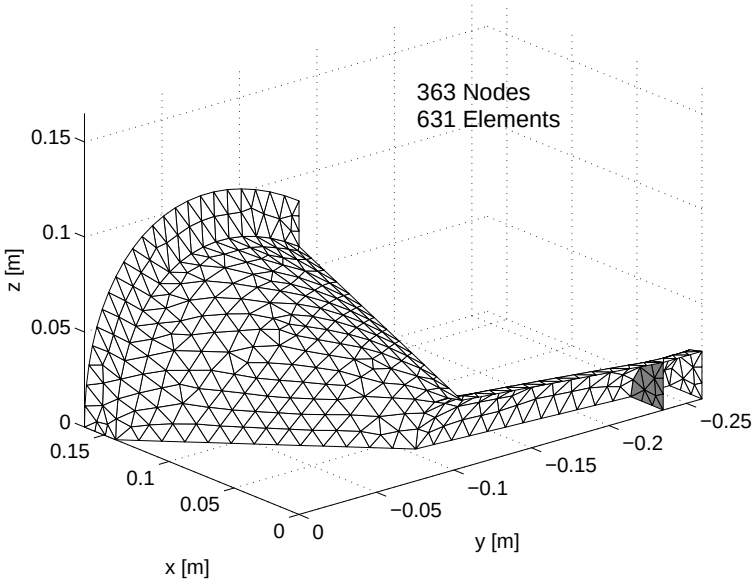


(b) Source superposition

Figure 5.4: Surface mesh of the exponential horn, 6 *elements per wavelength at* 5000 *Hz.*

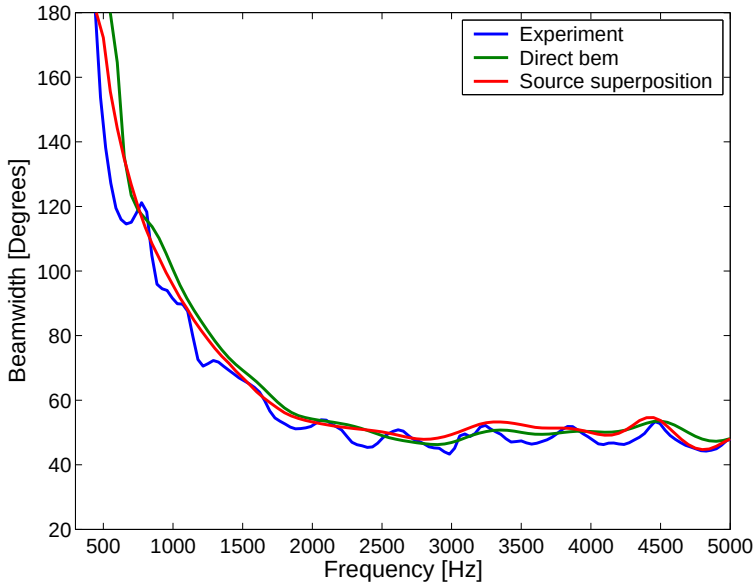


(a) Direct BEM

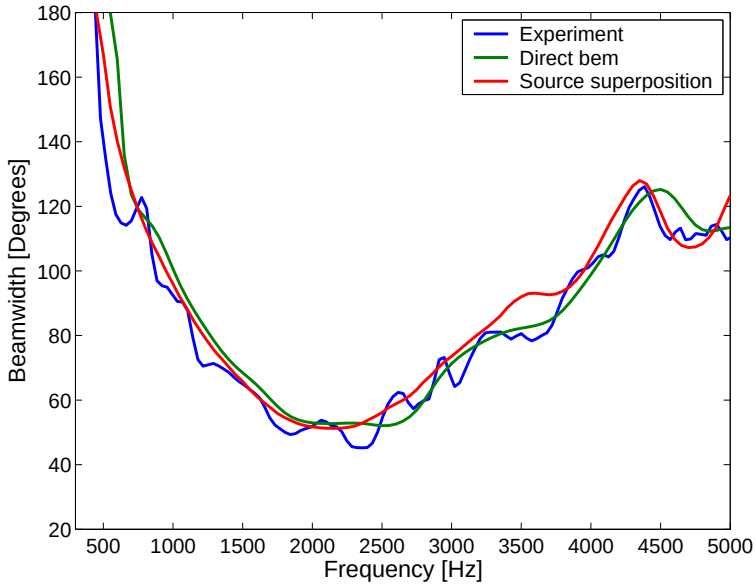


(b) Source superposition

Figure 5.5: Surface mesh of the conical horn, 6 *elements per wavelength at* 5000 *Hz.*



(a) Exponential



(b) Two step conical

Figure 5.6: Comparison of measured and calculated beamwidth for the direct BEM and source superposition techniques, 6 *elements per wavelength at* 5000 *Hz.*

5.4 Comparison at higher frequencies

The upper frequency considered in the previous section (5000 Hz) was limited by practical consideration of the computational time required for the direct BEM. Using the techniques developed in Chapter 4 for the source superposition technique, a more useful limit of 12000 Hz is investigated. This corresponds to a $ka \approx 36$, with a the radius of the horn mouth including the flange. Previous numerical studies of horn loaded loudspeakers that model beamwidth appear to have been limited in their upper frequency range to below $ka \approx 15$ (Section 2.3).

In this section, the source superposition technique is used to simulate the horns with a linear variation in frequency between 300 Hz and 12000 Hz. At frequencies higher than the cut on of the first mode at the throat (> 610 Hz, see Table 2.1), the assumption of a plane wave produced at the horn throat by the compression driver can be questioned, and experimental evidence of this has been found by Behler and Makarski (2003). The GMRES iterative solver with a loose (1×10^{-3}) tolerance (Section 4.4.2) is used for all simulations. Three different meshes are used, with various meshing strategies. Figure 5.7 shows a mesh for each horn at a nominal mesh density of 6 elements per wavelength, Figure 5.8 shows a 3 elements per wavelength mesh and Figure 5.9 shows a 3 elements per wavelength mesh, with 12 rotationally symmetric sectors, to take advantage of the increase in assembly speed available with this symmetry (Section 4.4.3). For shapes that are not rotationally symmetric, this last meshing strategy is not applicable. Horns used in the cinema industry are not generally rotationally symmetric (they are generally quarter symmetric), and the numerical methods used in this thesis are general and hence applicable to these horns. This is, in fact, a theme that runs through this thesis, that any optimisation method developed should be generally applicable, and is the reason an axisymmetric BEM was not used.

The results obtained show a general excellent agreement between the numerical results and the experimental results, as shown in Figure 5.10. The exponential horn shows good

agreement to above 10000 Hz, where the experimental and numerical methods diverge slightly. The results for the two step conical horn show excellent agreement over the entire frequency range.

The reason for the disagreement between results in the exponential horn above 10000 Hz is unclear. Figure 5.11 compares the measured and calculated directivity at a frequency of 11000 Hz, where there is some disagreement. Both numerical results give a slightly wider beamwidth than experiment. It is possible that higher order modes generated by the compression driver at these frequencies are able to propagate down the smooth variation in cross section of the exponential horn. This phenomenon would not be modelled by the numerical method, because only a plane wave is input at the mouth of the horn. This behaviour is not exhibited by the two-step conical horn, and it is hypothesised that the abrupt junction between the steps is already generating higher order modes, hence the good agreement between experimental and numerical results. The degree of agreement between the results is extremely good, given the resonance of the diaphragm in the compression driver is probably below 10000 Hz, and the difference in path length of the phasing plug (see Figure 2.4) is in the order of 3.4 cm.

The time taken to compute the solutions is shown in Table 5.3, showing the 3 elements per wavelength, rotationally symmetric method is very computationally efficient, whilst still giving sufficient accuracy for comparison to experimental results, at least to a frequency of 10000 Hz, above which deficiencies in modelling boundary conditions at the horn throat may become a factor.

The total time required for simulation of these horns using the fastest method is still about 17 minutes. This time is a significant saving over the estimated 16.5 hours required by the direct BEM for this same frequency range, however it is still too long for inclusion in an optimisation method.

The frequency resolution of these simulation is quite fine, with a 50 Hz frequency spacing. This can be reduced whilst still retaining information about the performance of the horn.

Method	Total [seconds]	Time / Freq [seconds]	Factor
6EPW	14172	60	15
3EPW	1537	7	1.8
3EPW, rotational symmetry	992	4	1

(a) Exponential

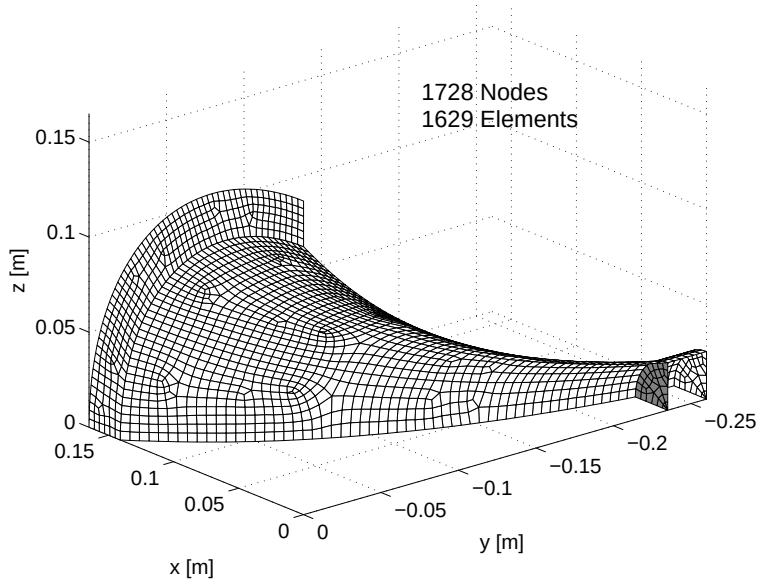
Method	Total [seconds]	Time / Freq [seconds]	Factor
6EPW	14870	63	15.8
3EPW	1908	8	2
3EPW, rotational symmetry	1056	4	1

(b) Two step conical

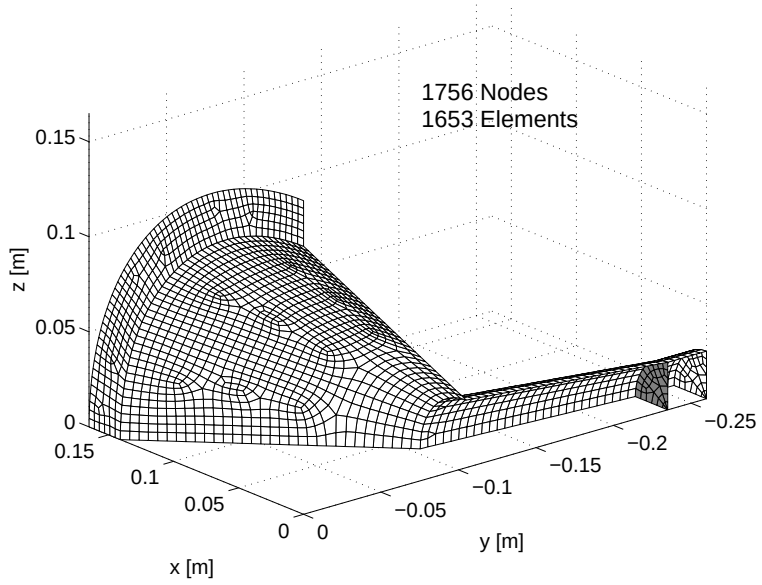
Table 5.3: Comparison of computational time taken for different meshing strategies.

Figure 5.12 shows the results at different frequency resolutions. The results at the lowest frequency resolution give the same broad trends as the finest resolution, whilst losing some of the fine details; however in comparison to the experimental results, the agreement is excellent. The estimated time required by the method at a linear frequency resolution of 400 Hz is 2 minutes, making this technique suitable for optimisation of these horns over this frequency range.

In summary, the source superposition technique, using a GMRES solver with a loose (1×10^{-3}) tolerance, 3 elements per wavelength, reducing assembly time by taking advantage of rotational symmetry and using a linear frequency spacing of 400 Hz can significantly reduce the time required to simulate horn loaded loudspeakers. It gives results for beamwidths that are suitable for design purposes, although results above 10000 Hz may not be accurate due to inadequacies of modelling the compression driver.

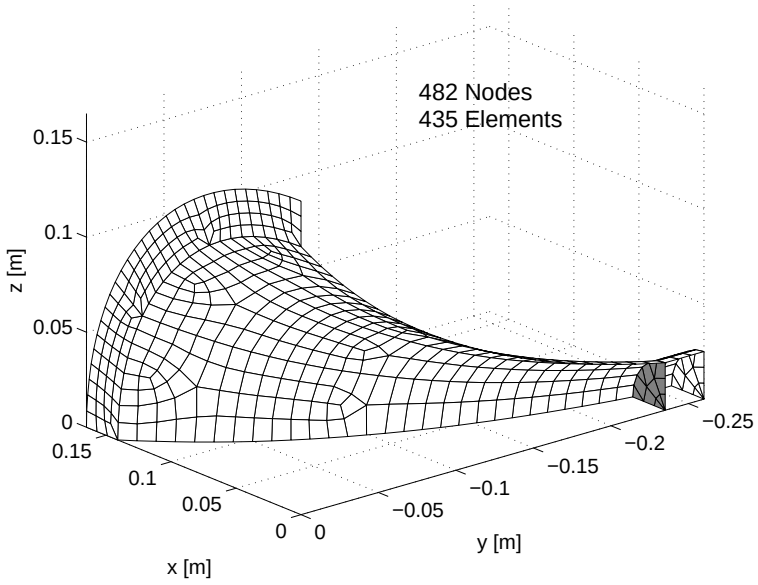


(a) Exponential

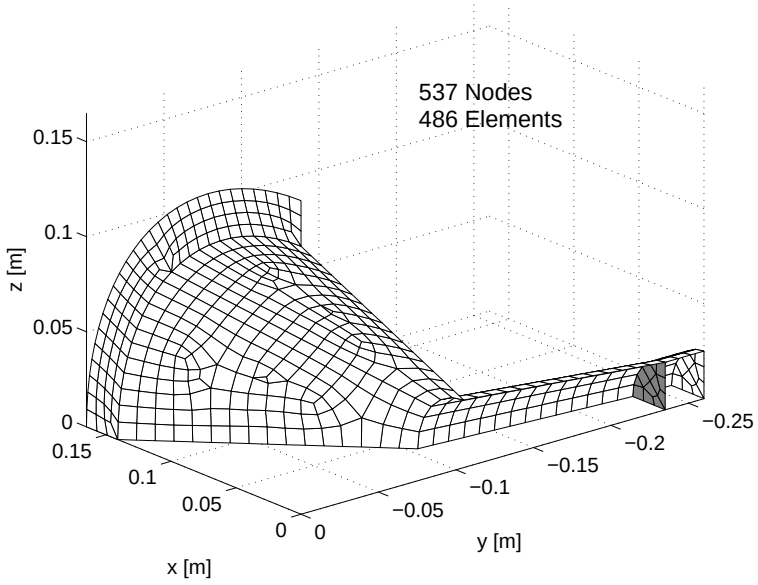


(b) Two step conical

Figure 5.7: Surface mesh of the exponential and conical horns, 6 *elements per wavelength at* 12000 *Hz.*

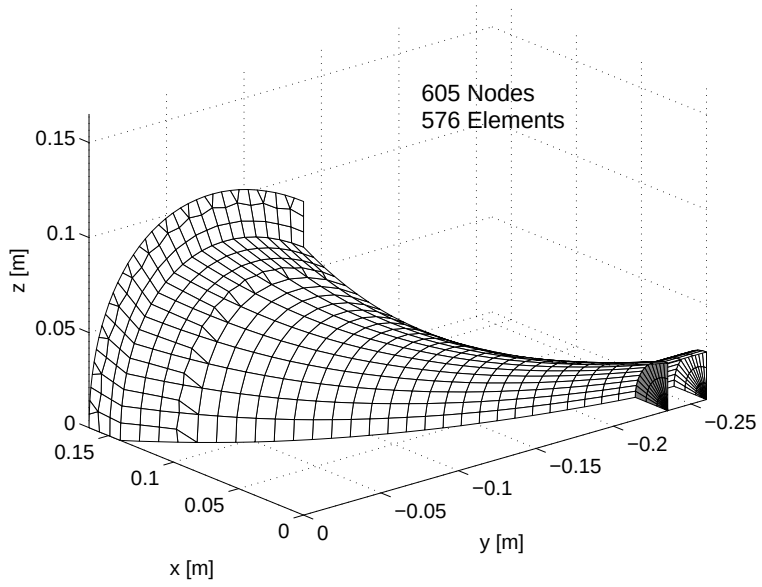


(a) Exponential

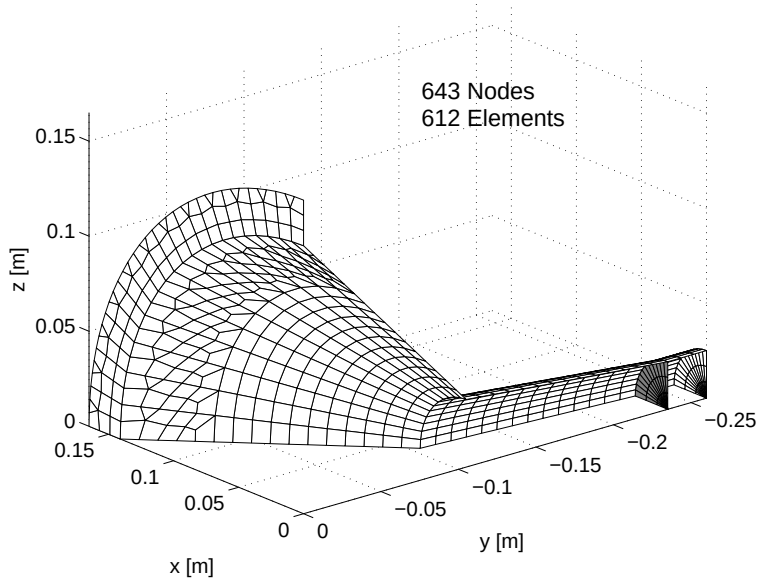


(b) Two step conical

Figure 5.8: Surface mesh of the exponential and two step conical horns, 3 *elements per wavelength at* 12000 *Hz.*

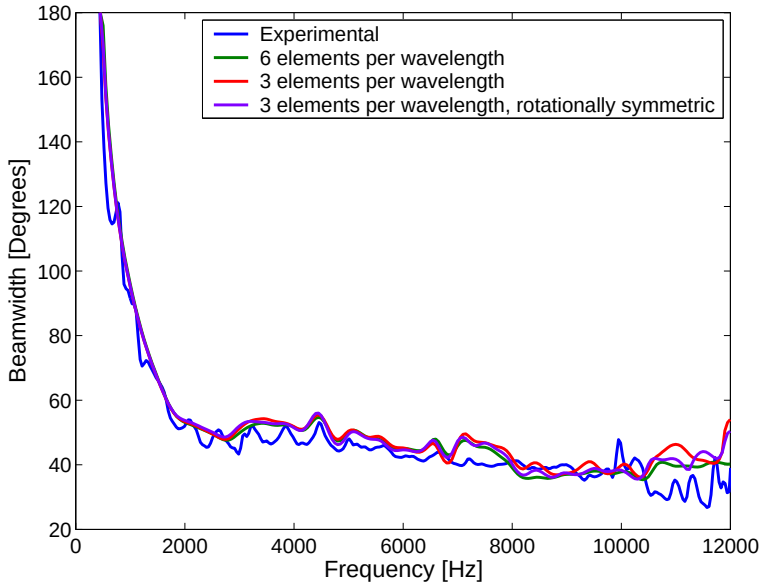


(a) Exponential

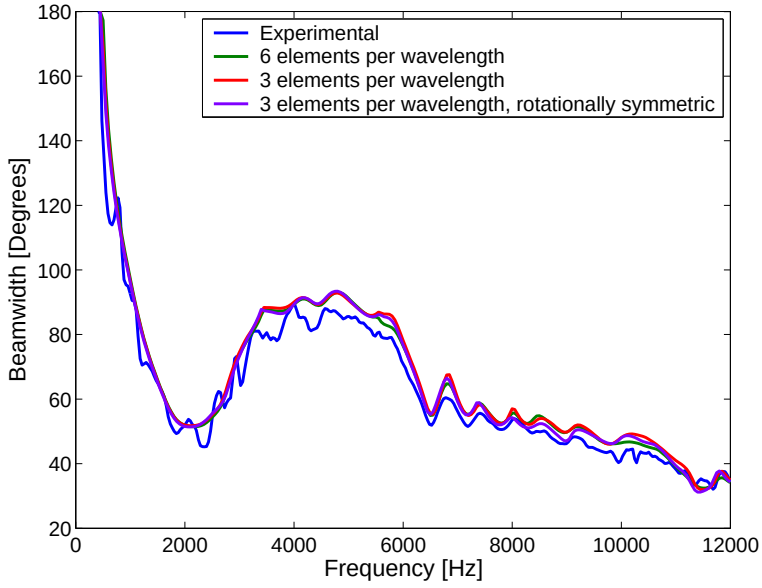


(b) Two step conical

Figure 5.9: Surface mesh of the exponential and conical horns, 3 *elements per wavelength at* 12000 *Hz with* 12 *rotationally symmetric sections.*



(a) Exponential



(b) Two step conical

Figure 5.10: Comparison of measured and calculated beamwidth for the source superposition technique with different meshing strategies.

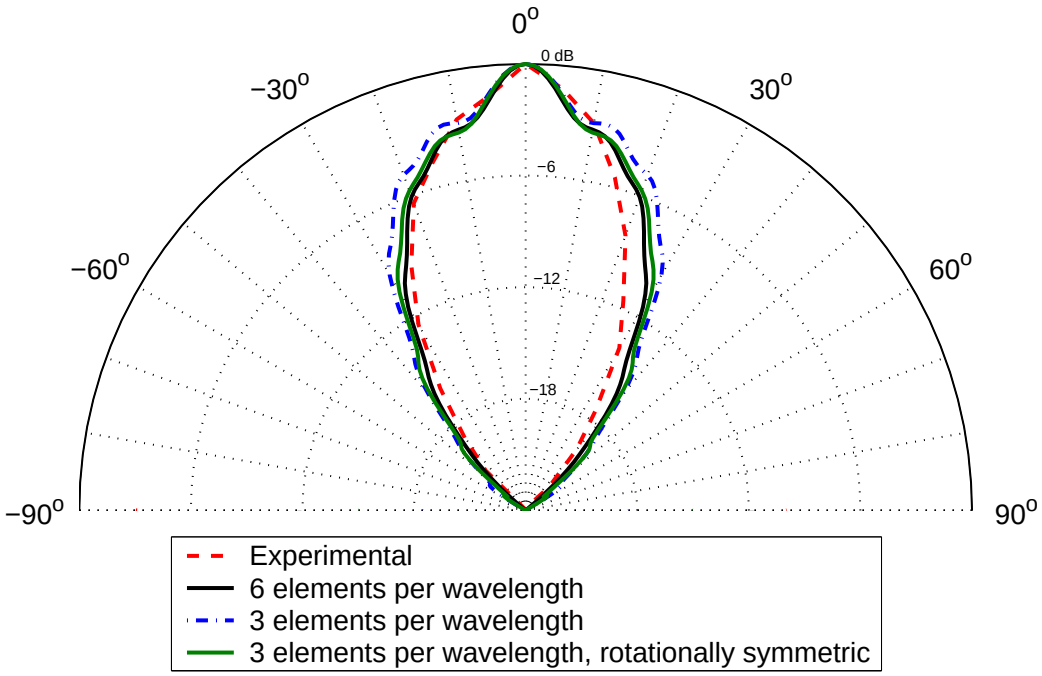
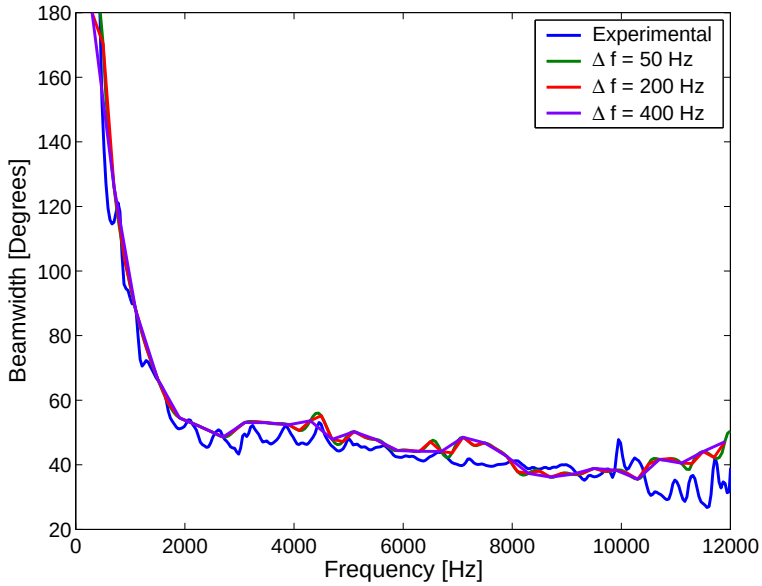
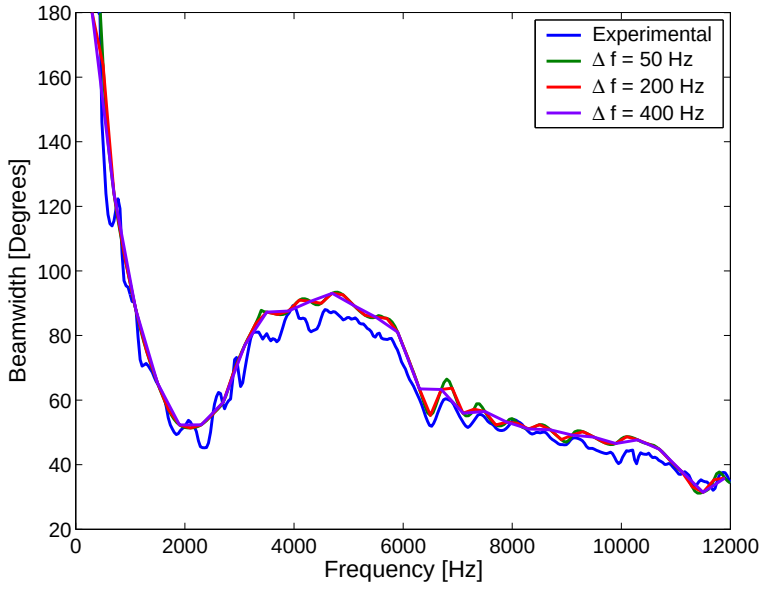


Figure 5.11: Comparison of measured and calculated directivity at 11000 *Hz.*



(a) Exponential



(b) Two step conical

Figure 5.12: Comparison of measured and calculated beamwidth for the source superposition technique with different frequency resolutions.

5.5 Conclusions

The aim of this chapter was to validate the source superposition numerical model by comparison to experiment, and to alternative numerical methods, such as the direct BEM. It has been shown that such models are capable of reproducing the sound field generated by horn loaded loudspeakers from a specification of the horn geometry. The accuracy of the reproduction was found to be adequate for design purposes within the given frequency range. Both the direct BEM and the standard source superposition technique are capable of reproducing the experimental beamwidth, however the source superposition technique is considerably faster.

Techniques to speed up the solution times the source superposition technique, reported in Chapter 4, have been applied to the modelling of horn loaded loudspeakers. It was found that using:

- a GMRES solver with a loose (1×10^{-3}) tolerance;
- a mesh with 3 elements per wavelength;
- reducing assembly time by taking advantage of rotational symmetry;
- a linear frequency spacing of 400 Hz

can significantly reduce the time required to simulate horn loaded loudspeakers. It gives results for beamwidth that are suitable for design purposes up to frequencies of about 10000 Hz for these horns. These numerical models are now suitable for use in optimisation techniques to be developed in further chapters.

Chapter 6

Frequency independent beamwidth transducers

The “holy grail” of horn design is twofold; an easily specified frequency independent beamwidth, and a smooth frequency response. Work performed on Constant Beamwidth Transducers (CBT) for sonar applications both theoretically (Rogers and Van Buren, 1978) and experimentally (Van Buren et al., 1983) has both of these desirable features.

The concept used in the development of CBTs is explored in this chapter in relation to horn design, giving cues as to how to design an optimum horn. An efficient semi-analytical model of the transducer is used to develop methods for robust beamwidth optimisation.

6.1 Introduction

The concept of a Constant Beamwidth Transducer (CBT) was suggested by Rogers and Van Buren (1978) with the primary aim of developing a sonar or ultrasonic projector that has a theoretically unlimited frequency independent beamwidth. The uses suggested

for the transducer included broadband echo ranging, high data rate communications and highly directive ultrasonic transducers. The claimed advantages for CBTs are extensive; uniform acoustic loading, extremely low side-lobes, virtually no near-field, and an essentially constant beam pattern for all frequencies above a certain cutoff frequency. The CBT concept is not just theoretical, as a comprehensive paper was published (Van Buren et al., 1983) on the design and implementation of such a device. An experimental CBT sonar projector was constructed and results presented over a wide range of frequencies, exemplifying the CBT advantages.

Further evidence that the CBT theory could be fruitful in the design of horns lies with the work of Keele (2000), who brought CBT theory to the notice of the audio community by applying it to arrays of discrete sources on a spherical cap. The theory has been applied to two dimensional arrays, such as those used in stadium sound (Keele, 2003), and also to flat panel arrays by the use of signal delays (Keele, 2002). This flat panel implementation is of interest to horn designers because it shows that the desired velocity profile can be achieved by shapes other than spherical surfaces.

Of very great interest to practical horn designs is the ability to independently control the two axes of beamwidth. Keele (2000) shows evidence that this is possible with CBT theory by showing the results of a calculation of a shaded array on the surface of a toroid, giving the desired beamwidth control in each axis. A more rigorous solution could be derived using spheroidal (Falloon et al., 2003) or ellipsoidal functions (Abramov et al., 1995, Levitina, 1995) although that approach would not be a trivial exercise.

The work described in this chapter aims to examine the applicability of the concept used in the development of constant beamwidth transducers to horn design using the numerical methods developed in previous chapters, and to develop optimisation techniques to be used in future chapters.

First, the theory of constant beamwidth transducers, which is based on the sound radiated by a specific velocity distribution on the surface of a sphere, is reviewed. The CBT

velocity distribution, as well as similar velocity distributions suggested by Jarzynski and Trott (1978) and Geddes (2002) is examined. This is done by extending the numerical technique developed in Section 4.2.1 to calculate arbitrary velocity distributions. Results calculated using the semi-analytical method are compared with results calculated using the source superposition technique (Section 4.2.4) to examine the effect of removing part of the spherical surface of the CBT. Then various optimisation techniques are utilised to find the optimum velocity distribution of a CBT. This is done as a precursor to finding an optimum horn geometry, which is discussed in future chapters. Finally conclusions as to the utility of CBT theory and its application in horn optimisation are drawn.

6.2 CBT Theory

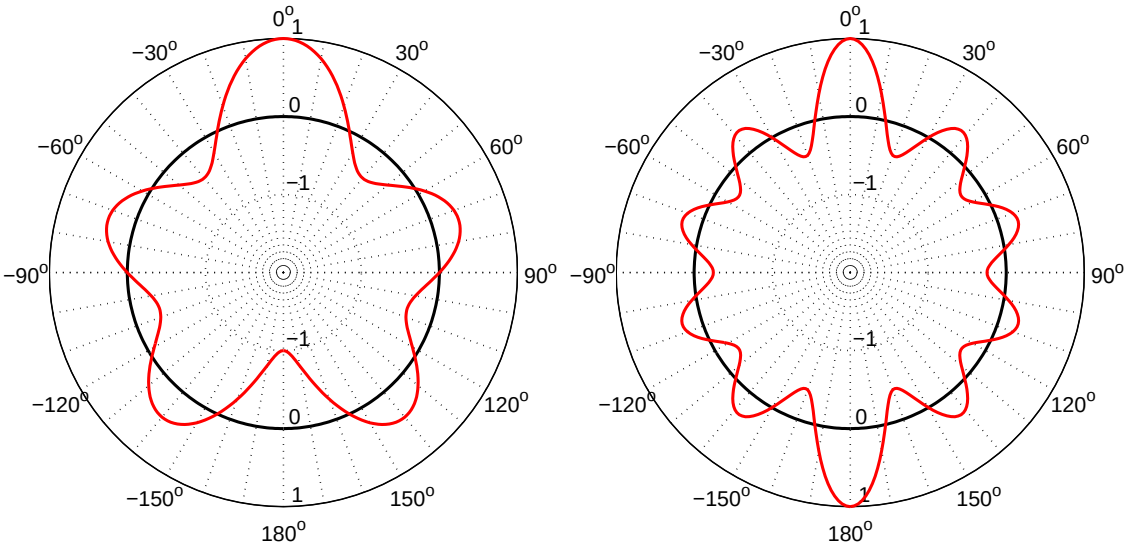
Acoustic radiation from the surface of a sphere can be calculated analytically using an infinite sum of orthogonal functions (Morse and Ingard, 1986). The derivation of the pressure field produced by an arbitrary velocity profile on the surface of a sphere (a velocity profile that varies with angular position, and given by $U(r,\theta)|_{r=a}$) was given in Section 4.2.1, with Equation 4.21 for pressure and Equation 4.19 for velocity coefficient reproduced here for reference as Equations 6.1 and 6.2.

$$p = -j\rho c \sum_{n=0}^{\infty} U_n P_n(\cos\theta) \frac{h_n^{(2)}(kr)}{h_n'^{(2)}(ka)} \tag{6.1}$$

$$U_n = \left(n+\frac{1}{2}\right) \int_0^{\pi} U(r,\theta)|_{r=a} \; P_n(\cos\theta)\sin\theta d\theta \tag{6.2}$$

CBT theory recognises that if the velocity profile, $U(r,\theta)|_{r=a}$, on the surface of the sphere is a single complete Legendre polynomial, P_n, then $U_{m\neq n} = 0$ due to the orthogonality of Legendre polynomials, and the pressure profile p contains only a single term in the summation. The angular variation in pressure is the same as the chosen velocity profile at all frequencies. In fact, any other velocity profile will lead to a frequency dependent

far-field response. Figure 6.1 shows the angular variation of two orders of Legendre polynomials.



(a) $P_5(Cos(\theta))$ (b) $P_{10}(Cos(\theta))$

Figure 6.1: Angular variation of Legendre polynomials of order (a) 5 and (b) 10.

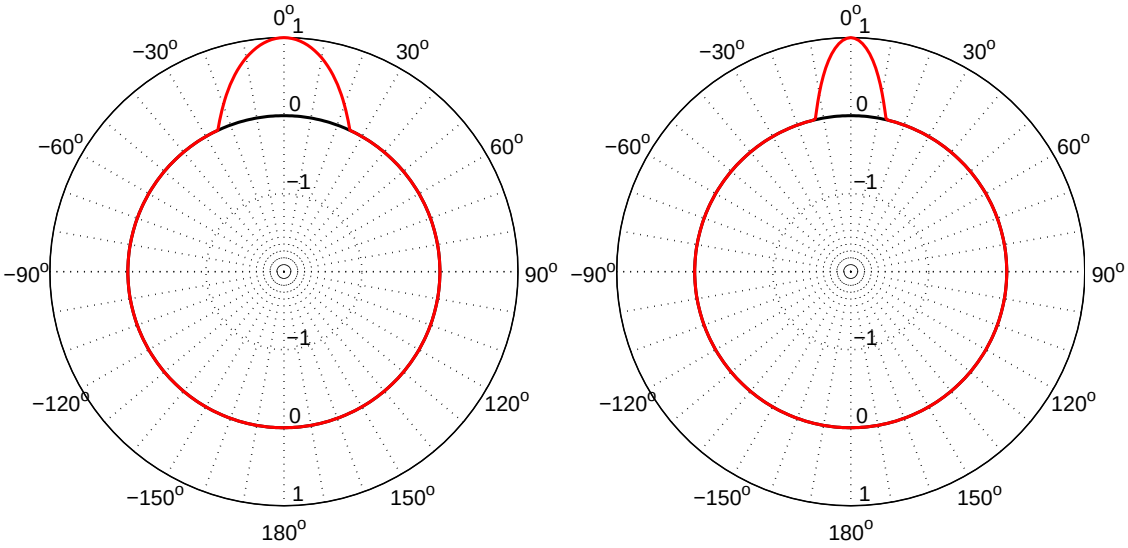
Using these profiles to produce a far-field pressure distribution is not desirable because of their lobed nature. Rogers and Van Buren (1978) have shown that by truncating the velocity profile at θ_0, the angle made by the first zero of the Legendre function[1] , i.e. $P_\nu(\cos(\theta_0)) = 0$, and setting the velocity to be zero for all angles greater than θ_0, an asymptotic form of frequency independent far-field response can be achieved.

The velocity profile for a CBT can be represented by Equation 6.3,

$$u_{cbt} = \begin{cases} u_0 P_\nu(\cos\theta) & , \quad 0 < \theta < \theta_0 \\ 0 & , \quad \theta_0 < \theta < \pi \end{cases} \tag{6.3}$$

where the order ν must be found for each θ_0. Figure 6.2 shows these responses for the same orders as Figure 6.1.

[1]It can now be a Legendre function of non-integer order (P_ν) rather a Legendre polynomial of integer order (P_n), because the restriction of single valued pressure around the sphere is relaxed.



(a) $P_5(Cos(\theta)), \theta < \theta_0; 0, \theta > \theta_c$ (b) $P_{10}(Cos(\theta)), \theta < \theta_0; 0, \theta > \theta_c$

Figure 6.2: Constant Beamwidth Transducer velocity profile for Legendre polynomial order (a) 5 and (b) 10.

This velocity profile was shown in Rogers and Van Buren (1978) to concentrate the energy in the lower order terms of the Legendre expansion, achieving the asymptotic (frequency independent) form quickly. This is the region of "geometric acoustics" as described by Morse and Ingard (1986, Page 340), and the ratio of surface pressure to applied velocity (specific acoustic radiation impedance) is $\rho_0 c$ over the entire surface of the sphere. One consequence of this, since the velocity is prescribed to be zero over the surface of the sphere at angles greater than θ_0, is that the physical surface of the sphere can be removed at these angles, leaving a spherical cap as the active part of the transducer. This reduces the size of the transducer, and gives hope to the design of horns using this theory by applying these boundary conditions over an imaginary spherical cap represented by the mouth of the horn. The problem then becomes how to shape the horn profile to give the required velocity distribution.

The velocity profile for a CBT in Equation 6.3 is cumbersome to use. Rogers and Van Buren (1978) give approximations for ν as a function of θ_0, and Keele (2000) gives a third order polynomial approximation to the velocity profile. This third order polynomial

approximation does not have the same slope as a Legendre polynomial ($\left.\frac{\partial u_{cbt}}{\partial\theta}\right|_{\theta=0} = 0$) at $\theta = 0$. Following the example of Keele, a fourth order polynomial approximation was found. Using boundary conditions $u_{cbt}(0) = u_0$, $u_{cbt}(\theta_0) = 0$ as well as $\left.\frac{\partial u_{cbt}}{\partial\theta}\right|_{\theta=0} = 0$ this approximation follows the same boundary conditions as the underlying Legendre function. The Legendre function P_ν was calculated using a hypergeometric series[2] in $\frac{1}{2}(1-\cos\theta)$ truncated at 100 terms, and the zeros found using the MATLAB function `fzero`. There was a small variation in the shape of the function when normalised by θ_0, over a range of θ_0 from 1°to 50°. A least squared fit between the normalised Legendre functions and the 4th order polynomial gives Equation 6.4.

$$u_{cbt} = \begin{cases} u_0\left(1 - 1.471\left(\frac{\theta}{\theta_0}\right)^2 + 0.1951\left(\frac{\theta}{\theta_0}\right)^3 + 0.2756\left(\frac{\theta}{\theta_0}\right)^4\right) & , \quad 0 < \theta < \theta_0 \\ 0 & , \quad \theta_0 < \theta < 180^\circ \end{cases} \tag{6.4}$$

The nominal design angle (θ_{nom}) is the angle at which the velocity profile reaches $0.5u_0$. For the CBT velocity profile, of $\theta_{nom} = 0.64\theta_0$.

6.3 Other methods for obtaining a frequency independent beamwidth

A review of the literature has found that methods other than those described in the previous section have been used to produce frequency independent beamwidths. The frequency independence of arrays has been addressed before (Chou, 1995, Van der Wal et al., 1996, and references contained within) by utilising a series of "nested sub-arrays". One array is optimised for a single frequency f_0 and the next is optimised for a frequency of $2f_0$. The size of the second array is half the first and the arrays are filtered so that the first is operating fully at frequency f_0 and is off at frequency $2f_0$, and vice versa for the second array.

[2]The hypergeometric series for P_ν can be written as $a_1 = -\nu$, $a_2 = \nu + 1$, $z = \frac{1}{2}(1-\cos\theta)$, and $P_\nu = 1 + a_1a_2z + \frac{1}{(2!)^2}a_1(a_1+1)a_2(a_2+1)z^2 + \frac{1}{(3!)^2}a_1(a_1+1)(a_1+2)a_2(a_2+1)(a_2+2)z^3 + \ldots$

Constant beamwidth is then achieved over an octave range. This type of constant beamwidth control would be very difficult to implement in a single horn, and is not considered further here.

Molloy (1968) describes a way to produce a prescribed far-field pressure pattern. The pressure profile is specified as a Chebychev polynomial, which has a single main lobe and many secondary lobes of the same height. The width of the main lobe and the ratio between the primary and secondary lobes can be specified independently. The surface velocity required to produce the far-field pressure can then be calculated. There is no comment in this paper as to the frequency dependence of the required surface velocity, and it is suspected that there is frequency dependence, unless the same asymptotic assumptions as the CBT can be made. It is not known how quickly the Chebychev polynomial reaches the asymptotic regime as compared with the CBT. It would also be difficult to specify the required side lobes when implemented as a horn, and again, this kind of transducer is not considered further.

The work of Jarzynski and Trott (1978) applies a similar reasoning to Section 6.2 in the design of a broadband constant beamwidth transducer array. They consider an acoustically transparent surface, presumably for the design of receiving arrays, but the velocity profile used should be equally applicable to a solid CBT. They use a velocity profile that is a linear combination of differing powers of $\cos^n\theta$,

$$u_{JT} = \frac{n}{2(2n+1)}\cos^n\theta + \frac{1}{2}\cos^{n+1}\theta + \frac{n+1}{2(2n+1)}\cos^{n+2}\theta \tag{6.5}$$

where n is an integer. If a non-integer n is used in Equation 6.5, u_{JT} becomes complex, and the absolute value should be used to describe the velocity variation. It was found that $n = 3.80$ would give a good approximation to $\theta_{nom} = 30^\circ$.

Geddes (2002, Section 6.6), in the context of horn (or waveguide) design, says that "it is possible, and reasonable, to do a waveguide design backward by specifying the desired

polar response pattern, calculating the required mouth velocity;...", and goes on to show some very impressive results (Geddes, 2002, Figure 6.21 - Optimized velocity profile and Geddes, 2002, Figure 6.22 - A Polar pattern with smooth angular variations) that show a wide band constant beamwidth. The velocity profile, which has presumably been calculated by some form of numerical optimisation, is similar to the Jarzynski and Trott profile. The Geddes velocity profile has been digitised from the figure, and fitted using least squares giving,

$$u_G = \begin{cases} u_0 \left(1 - \cos\left(\left(1 - \frac{\theta}{\theta_0}\right)^{1.16} \pi\right)\right) & , \quad 0 < \theta < \theta_0 \\ 0 & , \quad \theta_0 < \theta < \pi \end{cases} \tag{6.6}$$

and $\theta_{nom} = 0.45\theta_0$. The main difference between the CBT and the Geddes / Jarzynski and Trott profiles is the smooth decay as the profile approaches θ_0 (i.e. $\frac{\partial u}{\partial \theta}\Big|_{\theta=\theta_0} = 0$).

Figure 6.4 shows the uniformly vibrating cap, CBT, Geddes and Jarzynski and Trott velocity profiles, all for the same nominal beamwidth (60°, or a 30° half-angle). By definition, the value of the velocity profile at the half angle is one half of the value on the axis (or -6 dB down in SPL). It can be seen that the CBT, Geddes and Jarzynski and Trott velocity profiles extend beyond the nominal beamwidth angle due to their gradually decaying profiles. It is interesting to note that the first half of the Geddes and Jarzynski and Trott profiles are remarkably similar to the CBT profile, with much smoother "tails".

6.4 Semi-analytical calculation technique

Simulations using Equation 6.2 can be performed easily if the term U_n (Equation 6.1) can be calculated. In Section 4.2.1 these terms have been calculated analytically for a uniformly vibrating spherical cap on the surface of a sphere (Equation 4.23), where the

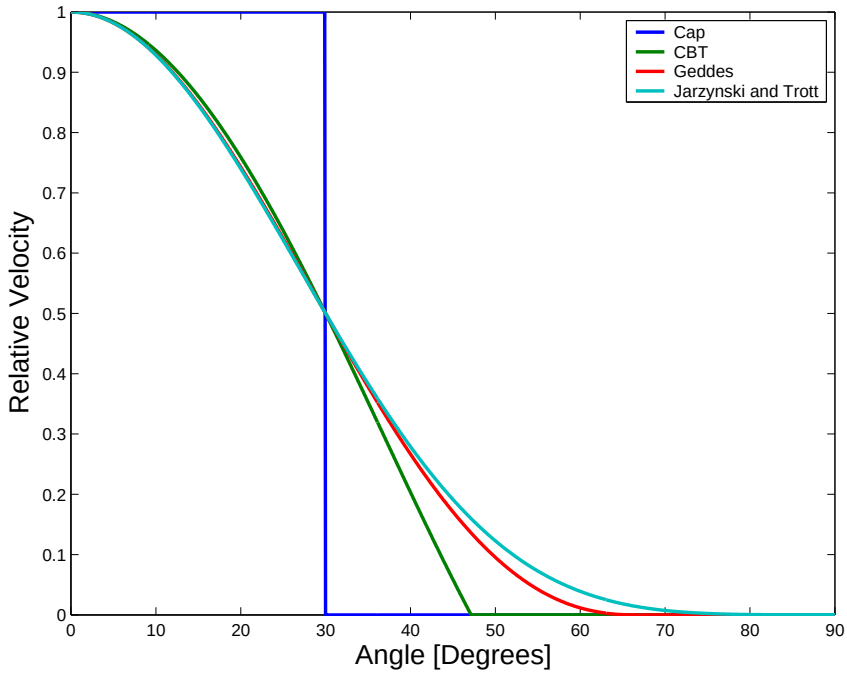


Figure 6.3: Comparison of velocity profiles for the same nominal half-angle (30°).

velocity of the vibrating cap is given by Equation 4.22, reproduced below as Equation 6.7.

$$u_C = \begin{cases} u_0 & , \quad 0 < \theta < \theta_0 \\ 0 & , \quad \theta_0 < \theta < \pi \end{cases} \tag{6.7}$$

Other velocity profiles do not have analytical solutions, and numerical techniques must be used to integrate Equation 6.1. A robust numerical integration routine, capable of integrating up to at least order $n = 400$, has been written to calculate these values for velocity profiles that may vary arbitrarily with angle. As the numerical integration can be computationally expensive, an adaptive routine is used to calculate terms in the series to a given tolerance, and a caching method used to avoid recalculation of expensive functions.

As a check of the semi-analytical numerical integration technique, a comparison between the beamwidth calculated using the semi-analytical method and that calculated using the analytical method described in Equation 4.23 was made. The beamwidth is defined as the "angle formed by the -6dB points (referred to the on-axis reading) and the source center"

(Davis and Davis, 1997), and is a measure of the distribution of sound in the specified plane (see Section 2.1.2). The calculations were undertaken for frequencies ranging from 300 to 20000 Hz for a 30° spherical cap with a uniform velocity distribution mounted on the surface of a sphere with radius 0.165 m. The general size and frequency range are similar to those of the horn loaded loudspeakers analysed in this thesis. Figure 6.4 shows no difference between the two solutions.

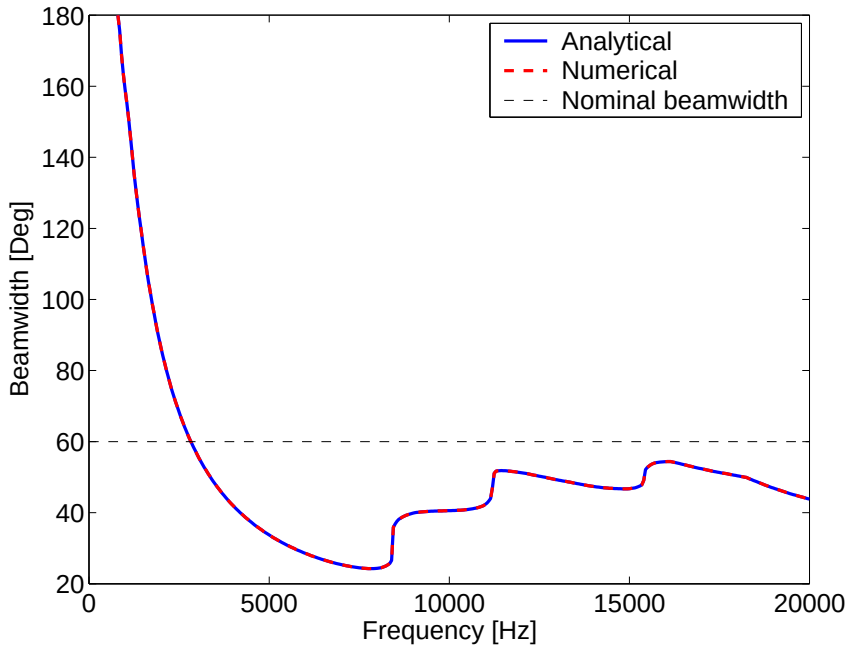


Figure 6.4: Comparison of analytical and numerical beamwidth calculations for 30° *spherical cap on the surface of a sphere with a uniform velocity distribution.*

At very high frequencies, the angular variation in the normalised pressure field should have the same angular variation as the underlying velocity field (Morse and Ingard, 1986, Page 340). In other words the beamwidth at high frequencies should have the same underlying angular variation as the prescribed velocity profile. This is not the case for the 30° spherical cap mounted on the surface of the sphere (which should have a 60°beamwidth), even at 20000 Hz (a *ka* of over 60). The reason behind this is the slow decay in magnitude of the velocity coefficients with increasing order (see Figure 6.7 for a comparison of velocity coefficient magnitude for different velocity profiles, and the discussion in Rogers and Van Buren (1978) and Van Buren et al. (1983)).

In summary this section shows that the semi-analytical calculation technique developed for calculating velocity coefficients (Equation 6.1) gives the same results for beamwidth as an analytical solution (Equation 4.23), and can be used for calculations where the velocity profile has an arbitrary angular variation.

6.5 Full sphere simulations

Simulations of the beamwidth produced by the four different velocity profiles shown in Figure 6.3 are compared using the semi-analytical method. To provide a fair comparison between each profile, the overall width of the spherical cap, b, was fixed at 0.330 m, the same dimension as used in the experimental horns in Chapter 3. For the same nominal half-angle, in this case 30°, each profile has a different finishing angle (θ_0). The radius of curvature of the each sphere, a, was calculated by $a = b/2\sin\theta_0$. The spherical cap profiles are illustrated in Figure 6.5 and the spherical radii, a, given in Table 6.1. It should be noted that the numerical integration technique will integrate over the entire sphere, however for most profiles the velocity is zero outside the cap. The Jarzynski and Trott profile is designed to fully cover the surface of the sphere, and never reaches zero. For this case, the radius of curvature is chosen to be the same as that of the Geddes profile, and a direct comparison of the change in velocity profile "tail" can be made.

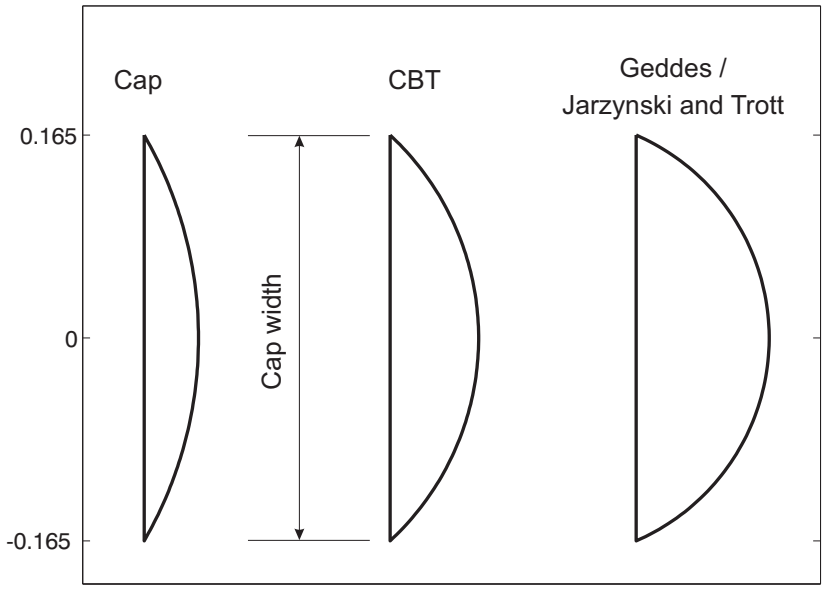


Figure 6.5: Comparison of spherical cap profiles for a constant cap width.

Calculations of the beamwidth generated by each profile were performed using the semi-analytical integration scheme up to 20000 Hz, and appear in Figure 6.6 in both linear

Profile	Radius a [m]
Piston	0.330
CBT	0.225
Geddes	0.180
Jarzynski and Trott	0.180

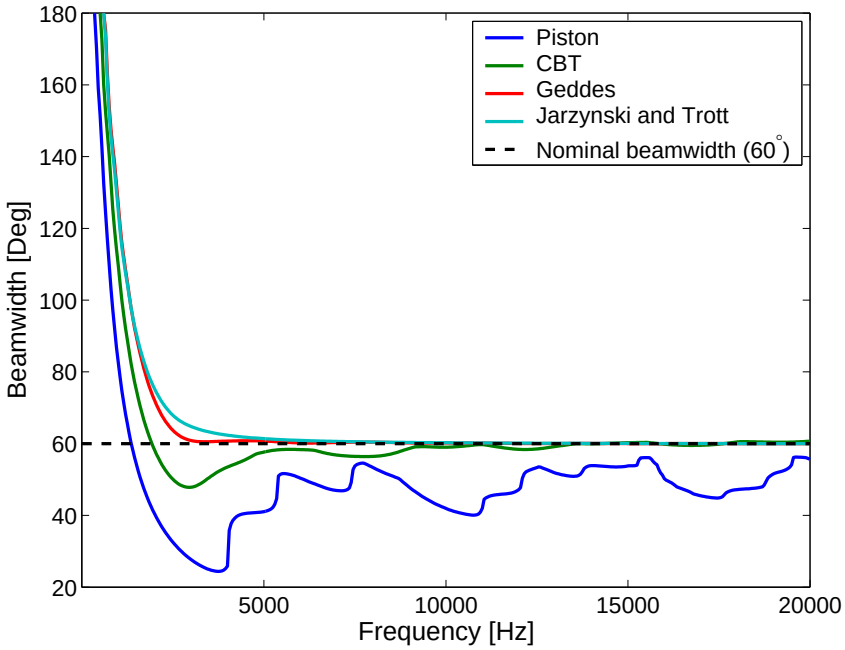
Table 6.1: Radius of curvature to keep a constant cap width.

and logarithmic form. The piston beamwidth does not reach the nominal angle as the frequency increases. The CBT beamwidth overshoots then approaches from below with rising frequency and reaches the nominal angle at high frequencies. The Geddes and Jarzynski and Trott profiles smoothly approach the nominal angle from above at low frequencies, and are superior in constant beamwidth performance to the CBT profile. The Geddes profile reaches the nominal beamwidth at a lower frequency than the Jarzynski and Trott profile. The low frequency performance of each velocity profile is different. The frequency at which the beamwidth starts narrowing is lowest for the piston profile ($\sim$ 1400 Hz), next highest for the CBT profile ($\sim$ 2000 Hz) and highest for the Geddes ($\sim$ 3000 Hz) and Jarzynski and Trott ($\sim$ 5000 Hz) profiles. This implies that there may be a trade off in low frequency performance for smooth beamwidth, and that the size of the velocity profile "tail" also has an affect on low frequency performance.

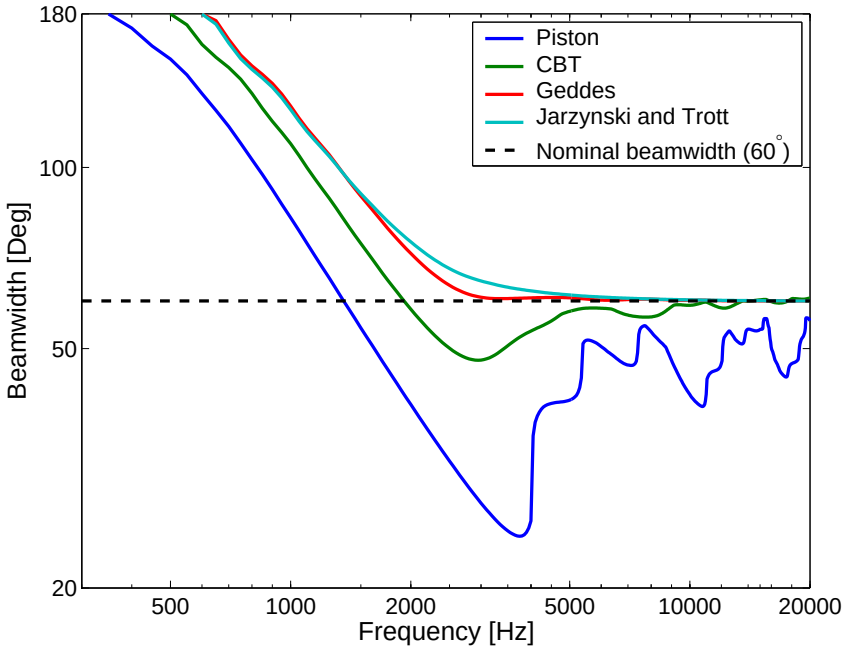
Figure 6.7 shows the rate of decay of the Legendre mode strength (U_n^2) with increasing order n, calculated using the semi-analytical integration technique. The Geddes profile decays most rapidly, followed by the CBT, then the piston. This shows that the energy is contained in the lower order terms of the Legendre expansion, achieving asymptotic (frequency independent) form with a rapid decay of the higher terms (Rogers and Van Buren, 1978). The Jarzynski and Trott profile initially rapidly decays to levels below the Geddes profile, but then the rate of decay decreases. This is possibly due to the profile not decaying to zero velocity at extreme angles, or more likely due to a "non optimal" velocity profile.

The Jarzynski and Trott profile is not further considered in this thesis because the Geddes profile is able to provide superior performance, and the specification of the Jarzynski and Trott profile using Equation 6.5 is difficult, as new values of n are required for each nominal design angle. It may be possible to analytically find a velocity profile that is "more optimal" than the Geddes profile but that is not considered in this thesis; instead robust numerical optimisation methods are needed to find optimum shapes of horn loaded loudspeakers, not velocity profiles over spherical caps.

This section compared the full sphere calculations of a spherical cap on the surface of a sphere with CBT, Geddes and Jarzynski and Trott velocity profiles. The frequency independent beamwidth performance of each profile was assessed, with the Geddes profile giving superior smoothness, possibly at the expense of low frequency performance.



(a) Linear scale



(b) Logarithmic scale

Figure 6.6: Beamwidth comparison for different velocity profiles.

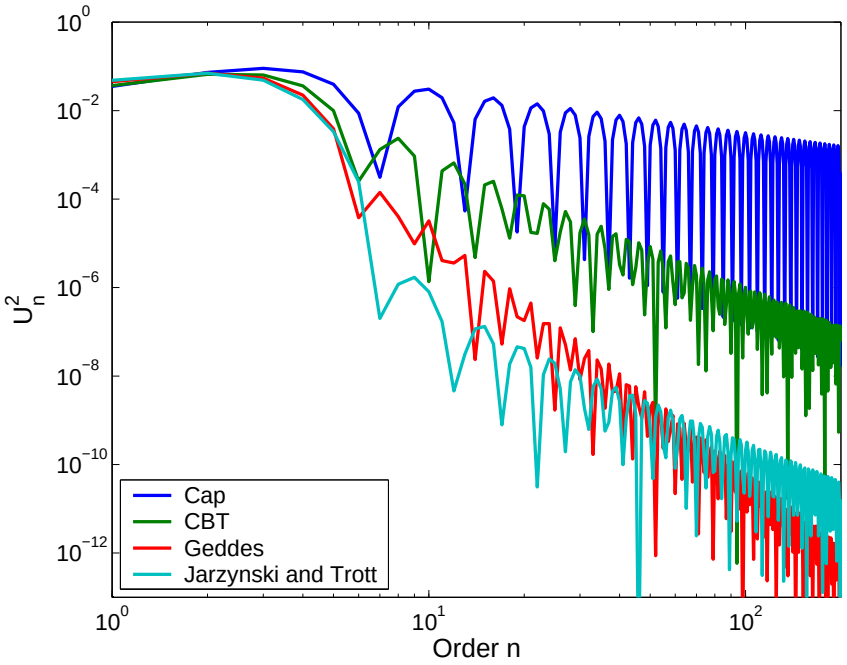


Figure 6.7: Legendre mode energy decay for different velocity profiles.

6.6 Spherical cap simulations

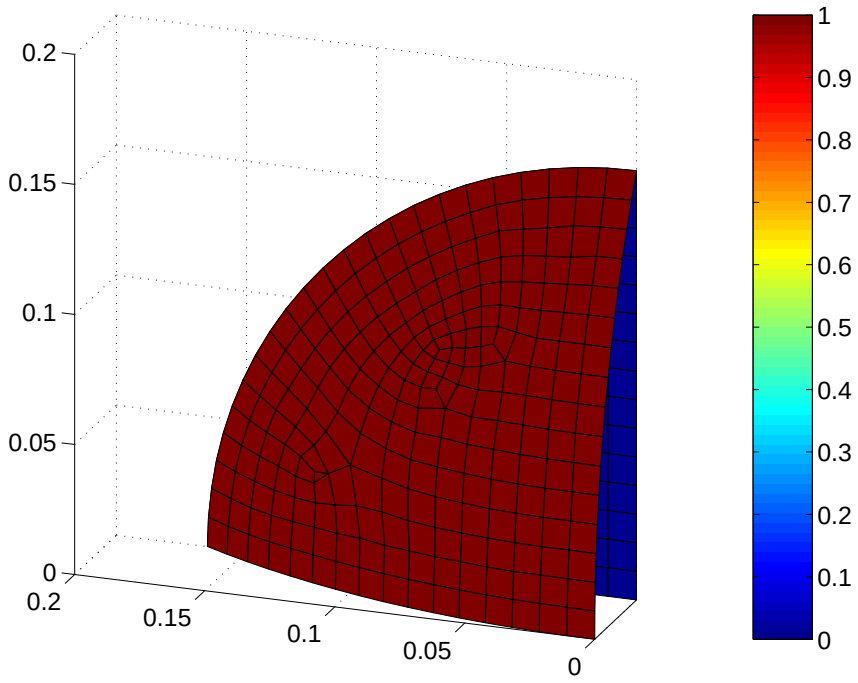
One of the noted advantages of a CBT is the ability to remove the surface of the sphere over which the velocity is zero. This cannot be tested using the semi-analytical integration technique (Section 6.4), as this implicitly assumes a solid surface over the entire sphere. The source superposition technique (Chapter 4.2.4) is a general numerical technique and can be applied in this situation. The source superposition technique was applied to simulating the uniformly vibrating spherical cap, CBT and Geddes velocity profiles, up to 12000 Hz in the absence of the part of the sphere that has zero prescribed velocity. A rigid plate covers the rear of the spherical cap to stop sound radiating from the back of the spherical surface. The mesh used for the spherical cap, coloured with normal velocity, appears in Figure 6.8 (a). In this case, the red surface represents the uniform velocity over the surface of a spherical cap, and the blue surface represents the plate covering the rear of the spherical cap, with zero velocity. A comparison of calculated beamwidth with the full sphere numerical simulation beamwidth appears in Figure 6.8 (b). There appears to be little difference in beamwidth over the entire frequency range considered.

The mesh used for the CBT velocity profile, coloured with normal velocity, appears in Figure 6.9 (a) and the calculated beamwidth compared to the full sphere numerical simulation beamwidth appears in Figure 6.9 (b). There is a small difference between the beamwidths at low frequencies, and virtually no difference over the rest of the frequency range considered.

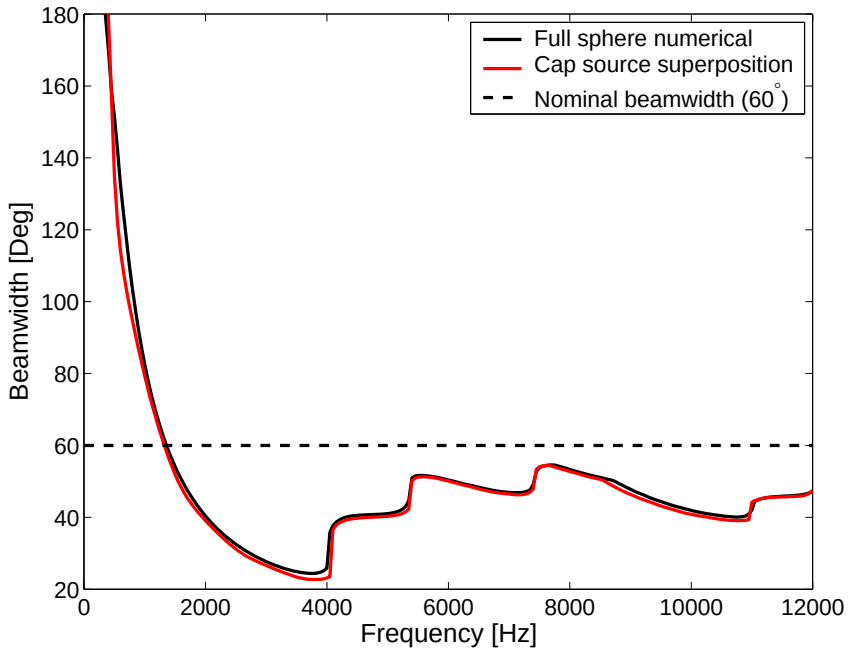
The mesh used for the Geddes velocity profile, coloured with normal velocity, appears in Figure 6.10 (a) and the calculated beamwidth compared to the full sphere numerical simulation beamwidth appears in Figure 6.10 (a). Again there is a small difference between the beamwidths at low frequencies, and virtually no difference over the rest of the frequency range considered.

This section has shown numerically that constant beamwidth theory does not apply ex-

clusively to spheres. As suggested in Rogers and Van Buren (1978), the surface over which the velocity is zero can essentially be removed with no loss of performance. This gives hope that a horn profile can be found that generates the same velocity profile over an imaginary spherical surface that covers the mouth of the horn, producing a frequency independent beamwidth above a certain limiting frequency.

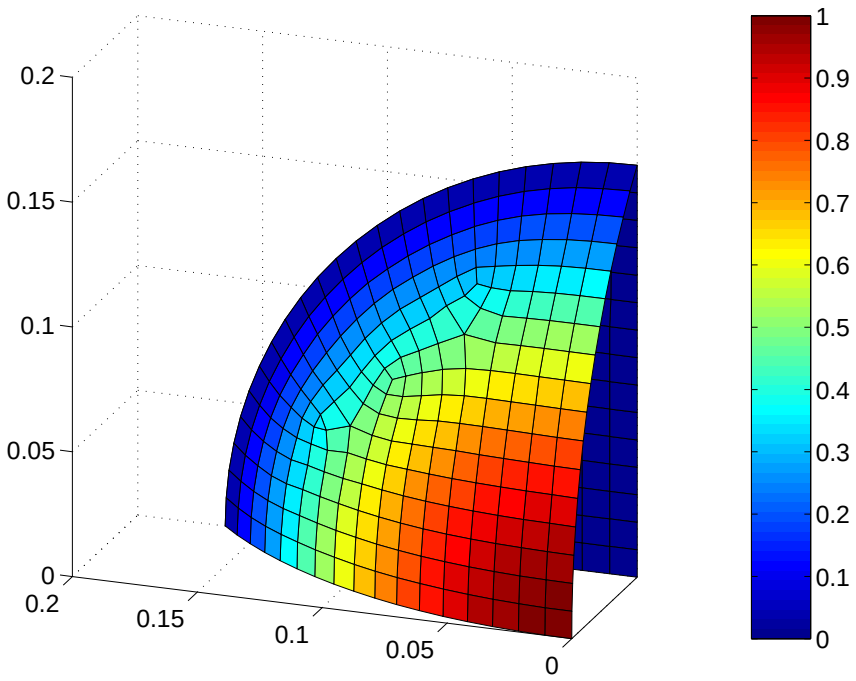


(a) Mesh

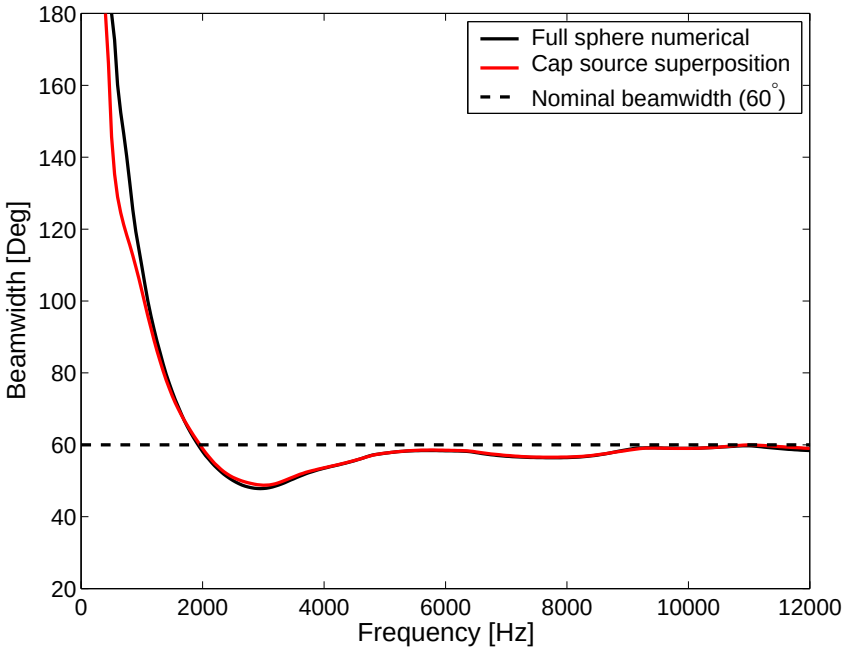


(b) Beamwidth

Figure 6.8: Comparison between the full sphere semi-analytical solution and the cap only source superposition technique for the spherical cap velocity profile.

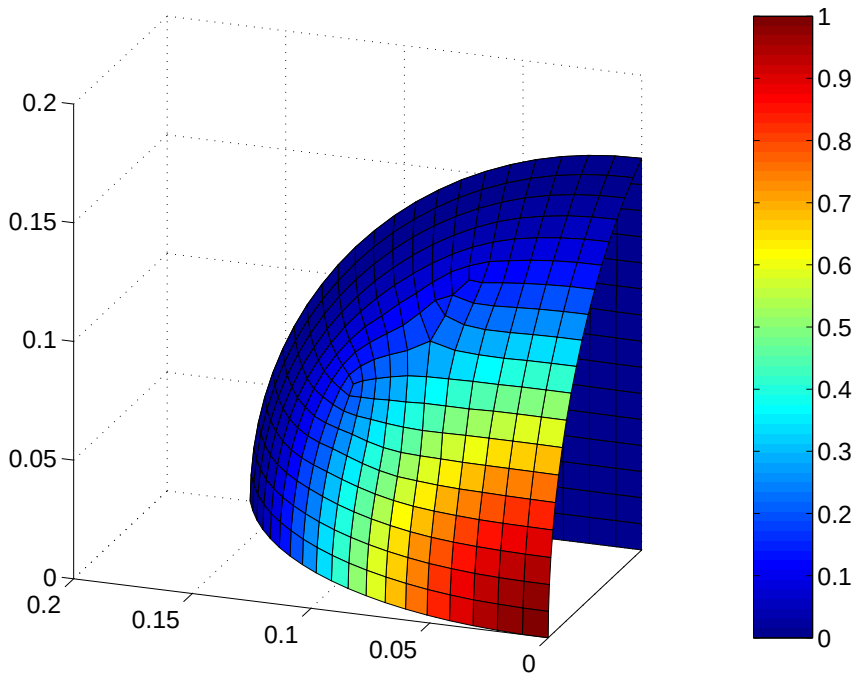


(a) Mesh

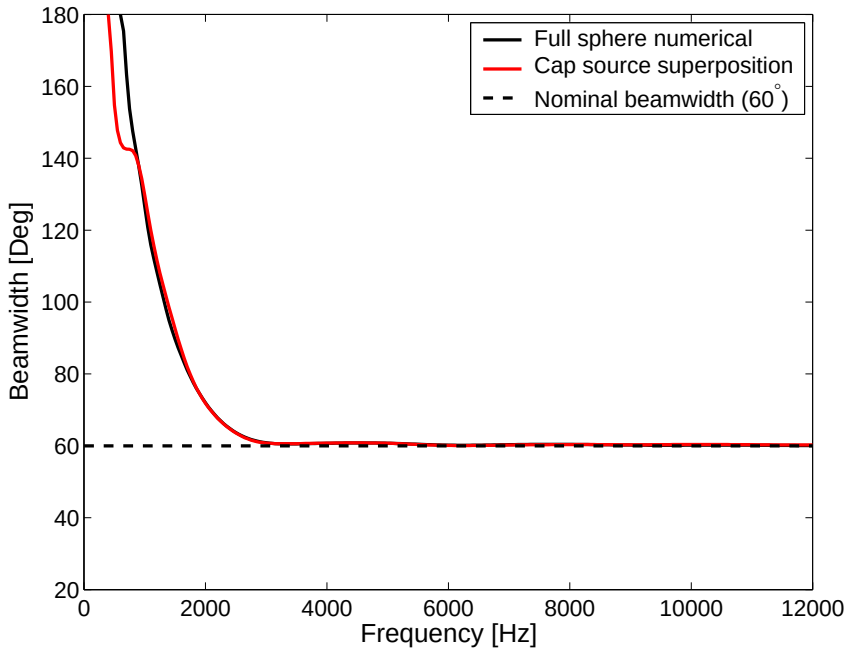


(b) Beamwidth

Figure 6.9: Comparison between the full sphere semi-analytical solution and the cap only source superposition technique for the CBT velocity profile.



(a) Mesh



(b) Beamwidth

Figure 6.10: Comparison between the full sphere semi-analytical solution and the cap only source superposition technique for the Geddes velocity profile.

6.7 Optimised Geddes velocity profile

From the previous sections it has been shown that the form of velocity profile described by Equation 6.6 has characteristics that allow an "optimal" beamwidth response. A generalised version of this profile can be described by,

$$u_{opt} = \begin{cases} u_0 \left(1 - \cos\left(\left(1 - \frac{\theta}{\theta_0}\right)^n \pi\right)\right) & , \quad 0 < \theta < \theta_0 \\ 0 & , \quad \theta_0 < \theta < \pi \end{cases} \tag{6.8}$$

where n is now a parameter that changes the shape of the profile as opposed to previously being a constant. This velocity profile now has two variables, n and θ_0.

Figure 6.11 shows the variation of the profile shape with the parameter n varied from 0.2 to 2, whilst holding nominal design angle (θ_{nom}, the angle at which the velocity profile reaches $0.5u_0$) constant. There is a large variation in normalised profile, from no tail for $n = 0.2$ to a very large tail for $n = 2$.

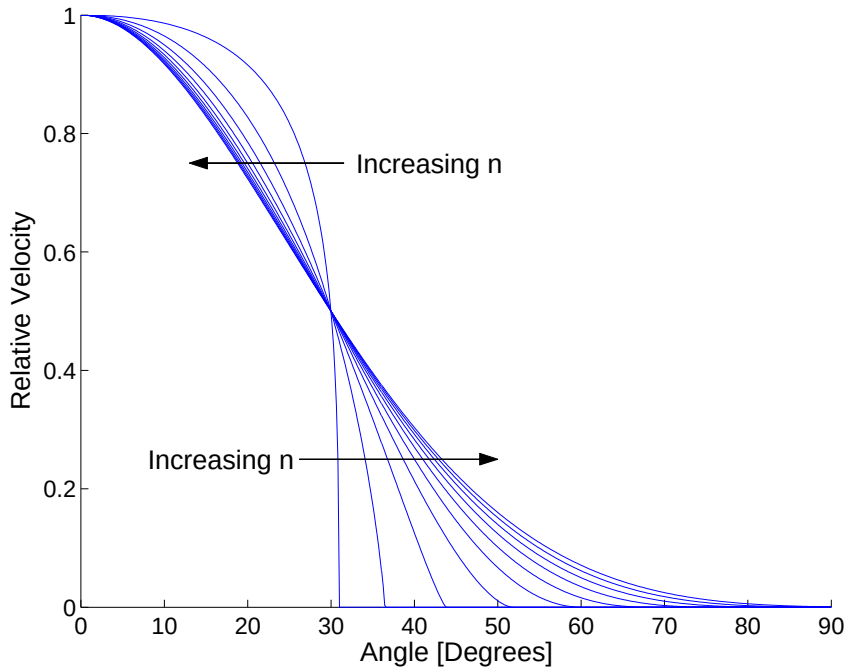


Figure 6.11: Comparison of velocity profiles as n changes from 0.2 to 2, θ_{nom} constant.

The obvious choice for optimisation of this profile, would be to set $\theta_{nom} = \mathcal{B}_{nom}$, the desired beamwidth, as we know that the far field pressure profile (and hence the beamwidth) will be proportional to the velocity profile in the asymptotic limit. Optimisation on the parameter n until a smoothness criterion is minimum would then ensure the asymptotic limit is reached at the lowest possible frequency. However as a more general test of the optimisation routines to be used in future chapters, both n and θ_0 will be varied.

6.7.1 Objective functions

For any optimisation, an objective function describing the relative merit of the current solution must be calculated. The stated objective for designing horn loaded loudspeakers is a smooth, frequency independent beamwidth. One type of objective function that is commonly used is a least squares objective function (Objective function #1),

$$\begin{gathered} \min \Phi_1 \\ \Phi_1 = \sum \left(\mathcal{B}\left(\mathbf{f} \geq f_{min}\right) - \mathcal{B}_{nom} \right)^2 \end{gathered} \tag{6.9}$$

where $\mathcal{B}(\mathbf{f})$ is a vector of beamwidths calculated using the semi-analytical numerical method over a range of frequencies described by the vector $\mathbf{f}$. The operator $\mathbf{f} > f_{min}$ selects only those frequencies above f_{min}, $\mathcal{B}_{nom}$ is the nominal (or desired) beamwidth. Examination of Figure 6.6 shows that for the "optimal" solution of Geddes, the beamwidth is smooth only above a certain limiting frequency (f_{min}), in this case approximately 3000 Hz.

The reason for selecting frequencies above a lower limit (f_{min}) is that at low frequencies the beamwidth is very wide. The inclusion of low frequency beamwidths would shift the mean value of the beamwidth up, and increase the standard deviation, and the resulting metrics would not represent "constant beamwidth behaviour". Currently the lower frequency limit is arbitrarily defined to be $f_{min} = 3100$, based on experimental evidence of the low frequency performance of horns with the same mouth size (Figure 5.10).

Objective function #1 is attractive because it concisely states the objective for designing horn loaded loudspeakers; that is, over the range of frequencies considered, any deviation away from the nominal beamwidth is penalised. Conversely, if all values of the beamwidth are equal to the nominal beamwidth, then the objective function is exactly zero. The objective function also does not require any constraints (other than perhaps an upper and lower bound), which may be an advantage in implementation of the optimisation algorithm.

To examine the potential performance this objective function, values of Φ_1 were calculated for a range of profile parameters, n varying from 0.2 to 2 and θ_0 from 20° to 90°. Each parameter was calculated at 25 evenly spaced points, for a total of 625 samples. In all cases, the diameter of the spherical cap on the surface of the sphere is held the same, so the curvature is allowed to change (as exemplified in Figure 6.5). The variation of Φ_1 is very large, and it may be preferable for both visualisation and optimisation to use the natural logarithm of Φ_1. This value is contoured in Figure 6.12, and shows multiple minima, pointed to by black arrows. The global minimum is represented as a red dot.

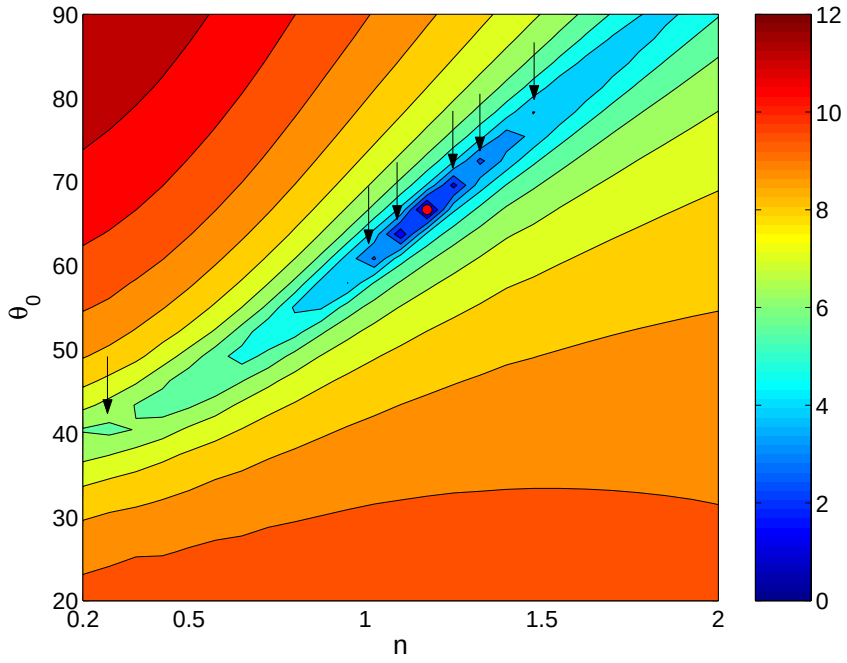


Figure 6.12: Contours of $\ln \Phi_1$*, the logarithmic least squared objective function.*

The presence of multiple local minima in the same vicinity is not helpful in finding a global minimum using gradient methods, and alternative objective functions are sought. In returning to the original stated objective for designing horn loaded loudspeakers; that is a "smooth, frequency independent beamwidth", the standard deviation of the beamwidth calculated over a desired frequency range is a measure of smoothness (the lower the better), and the mean value of the beamwidth calculated over a desired frequency range will give the target value for design ($\mathcal{B}_{nom}$). Some combination of the above measures may lead to a smoother objective function. Defining,

$$\Phi_2 = \texttt{mean}(\mathcal{B}(\mathbf{f} \geq f_{min})) \tag{6.10}$$

$$\Phi_3 = \texttt{std}(\mathcal{B}(\mathbf{f} \geq f_{min})) \tag{6.11}$$

where $\texttt{mean}(\mathbf{x})$ and $\texttt{std}(\mathbf{x})$ are the mean and standard deviation of vector $\mathbf{x}$, respectively. Figure 6.13 shows the variation of Φ_2 and Φ_3 over the same range of parameters used to generate Figure 6.12. There is a very smooth variation in Φ_2 and a relatively smooth "optimal basin" for Φ_3.

An ideal approach to finding an optimal solution would be a multi-objective optimisation (Belegundu and Chandrupatla, 1999, Deb, 2001) where both Φ_2 and Φ_3 would be optimised simultaneously to obtain a set of solutions that would be able to trade off smoothness for a mean beamwidth angle. The designer could then choose a solution from this set that was able to best meet their needs. However, full multi-objective optimisation is a topic of current research and beyond the scope of this thesis.

A simple form of multi-objective optimisation is to turn the multiple objective functions into a minimisation function and a secondary constraint. This is possible when there is a known (and achievable) target function value. In this case, the smoothness parameter (Φ_3) is used as the minimiser, and the mean beamwidth value (Φ_2) is used to constrain the function to the required value. This objective function (Objective function #2) can be

written as,

$$\min \Phi_3 \tag{6.12}$$

with equality constraint

$$\Phi_2 = \mathcal{B}_{nom} \tag{6.13}$$

In some cases, it may be necessary to relax the equality constraint to a bounded or pseudo-equality constraint such as,

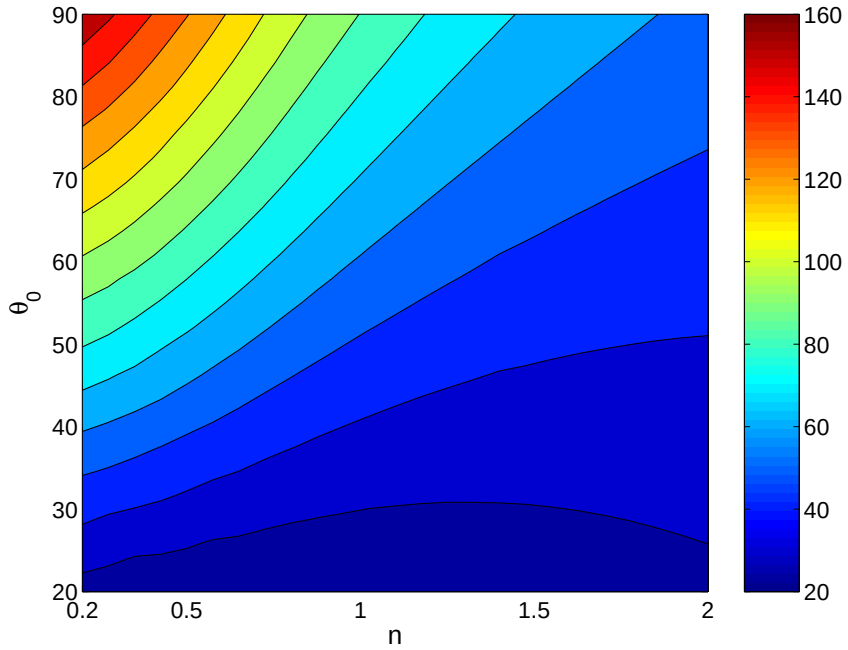
$$\mathcal{B}_{nom} - \varepsilon \leq \Phi_2 \leq \mathcal{B}_{nom} + \varepsilon \tag{6.14}$$

where ε is a tolerance parameter that describes how close to the nominal beamwidth the optimisation is constrained.

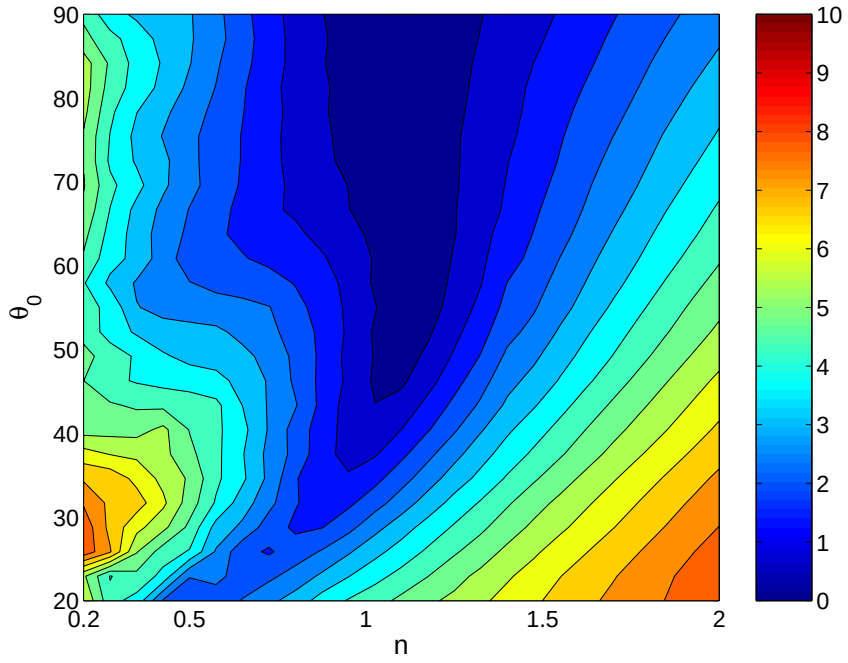
Equation 6.14 can be reformulated as two inequality constraints,

$$\begin{aligned} \mathcal{B}_{nom} - \Phi_2 &\leq \varepsilon \\ \Phi_2 - \mathcal{B}_{nom} &\leq \varepsilon \end{aligned} \tag{6.15}$$

Figure 6.14 shows the natural logarithm of the smoothness function, Φ_3, overlaid with a contour of $\Phi_2 = 60^\circ$ (the black line). The red dot shows the minimum of this function, and the black arrows show local minima of the smoothness function Φ_3. The intersection of the black line and the minimum value of the contoured function is a graphical solution to objective function #2 with $\mathcal{B}_{nom} = 60^\circ$. The use of the natural logarithm in this case is for visual acuity, as the range of variation of this function is not as severe as parameter Φ_1.



(a) Mean (Φ_1)



(b) Standard Deviation (Φ_2)

Figure 6.13: Contours of Φ_2 and Φ_3, the mean and standard deviation of the beamwidth.

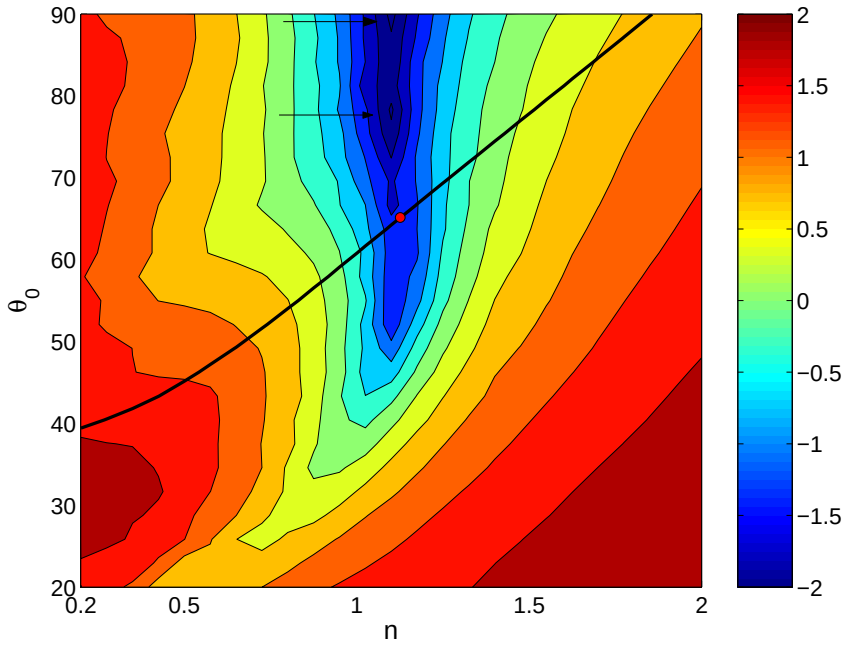


Figure 6.14: Contours of Φ_3, overlaid with with a contour of $\Phi_2 = 60°$.

6.7.2 Optimisation

There are many different optimisation techniques that can be used to minimise the objective functions given in Section 6.7.1. Belegundu and Chandrupatla (1999) give a good introduction to optimisation and Appendix B contains an overview of the optimisation techniques used in this thesis.

All optimisation methods evaluate the objective functions many times during the search for an optimal solution. Even though the semi-analytical numerical method is a relatively fast way to calculate the beamwidth, and hence functions Φ_1 and Φ_2, it is still time consuming. Because the solutions had already been calculated over a relatively fine rectangular grid to visualise the functions in Figures 6.12 and 6.13, an interpolation method called Kriging (See Appendix B.4) was used approximate the functions between the known values. This was done using the MATLAB DACE toolkit (Lophaven et al., 2002b,a). This technique effectively interpolates the function values and gradients between known values, and can be effective in reducing overall computational time when objective function or constraint function evaluations are expensive to calculate.

It was found that the MATLAB optimisation toolbox function `fmincon`, a Sequential Quadratic Programming (SQP, see Appendix B.2) optimisation technique, was able to find the optimal solutions for objective function #2 with equality constraints (Equation 6.13) for all initial starting positions tested. Depending on the starting position, between 30 to 60 function evaluations were required. SQP optimisation of objective function #2 with inequality constraints (Equation 6.15) would converge some of the time, depending on the constraint tolerance parameter ε and the initial starting position. The robustness of objective function #2 can be attributed to the fact that only a single minima is encountered when the optimisation is moves along the constrained path, even though function Φ_3 does contain local minima.

Finding the optimal solution for objective function #1 was more difficult due to the mul-

tiple local minima, and SQP would not converge robustly to the global minimum. A solution to the problem of robust convergence is to use a global optimisation technique, such as DIRECT (See Appendix B.3). This technique found a global minimum for both objective functions. It could robustly find a global minimum for objective function #1 with approximately 250 objective function evaluations. It was able to find the global minimum for objective function #2 with the constraint tolerance parameter $\varepsilon = 1$ in approximately 500 iterations. If the constraint tolerance parameter was set tighter, $\varepsilon = 0.1$, the solution was still changing after 1000 iterations, and the DIRECT method appears inefficient for pseudo-equality constrained optimisation. These results for both SQP and DIRECT are summarised in Table 6.2.

	Equality constrained SQP	DIRECT
OF #1	Would not robustly converge.	250
OF #2	$30-60$, depending on start point.	$500-1000$

Table 6.2: Number of objective function evaluations required for robust optimisation.

These results imply that the DIRECT global optimisation technique is best for solving the unconstrained minimisation of objective function #1, and that SQP would provide a robust optimisation technique for the equality constrained minimisation of objective function #2. The SQP algorithm performs well because, as can be seen from Figure 6.14, the minimisation problem along the constraint is a smooth function with a single local minimum. It has been found that the DIRECT algorithm's behaviour for objective function #2 is to rapidly find a solution close to the global minimum, then further iterations gradually improve on this approximation. For more general optimisation problems, it is suggested that a few iterations of the DIRECT algorithm with a loosely ($\varepsilon = 1$) inequality constrained objective function #2 will provide a good starting guess for the SQP algorithm, solving the equality constrained objective function #2. The number of DIRECT iterations that are needed will be problem dependent.

The previous optimisation results were obtained using interpolation methods (Kriging) to

obtain objective function values between those pre-calculated on a rectangular grid. This approach is untenable for general problems, especially in higher dimensions due to the large number of objective function evaluations requires. The interpolation method was used in this case to rapidly evaluate different objective functions. To examine the performance of these chosen algorithms on exact objective functions, the DIRECT solution for unconstrained minimisation of objective function #1 (as the SQP algorithm would not converge to a global minimum due to multiple local minima), and the SQP algorithm solving the equality constrained objective function #2, with a few initial DIRECT iterations, were calculated. The SQP algorithm used a finite difference approximation to the gradient, automatically calculated by MATLAB. The optimal values of the inputs, n and θ_0, along with the outputs, Φ_1, Φ_2 and Φ_3 and number of function evaluations (N) required appear in Table 6.3 for both objective function #1 and #2. The same parameters are given for the Geddes profile (Equation 6.6). These show that optimisations performed with both objective functions are able to find (marginally) better solutions than the original profile, with measures of the "smoothness" of the beamwidth, $\ln(\Phi_1)$ and Φ_3 having lower values than the Geddes profile, and the average value of the beamwidth, Φ_2 closer to the nominal design beamwidth.

	n	θ_0	$\ln(\Phi_1)$	Φ_2	Φ_3	N
Geddes	1.16	66.5	1.92	60.2	0.28	n/a
DIRECT, OF #1	1.12	65.1	0.75	60.0	0.22	1538
Equality constrained SQP, OF #2	1.13	65.2	0.83	60.0	0.22	61

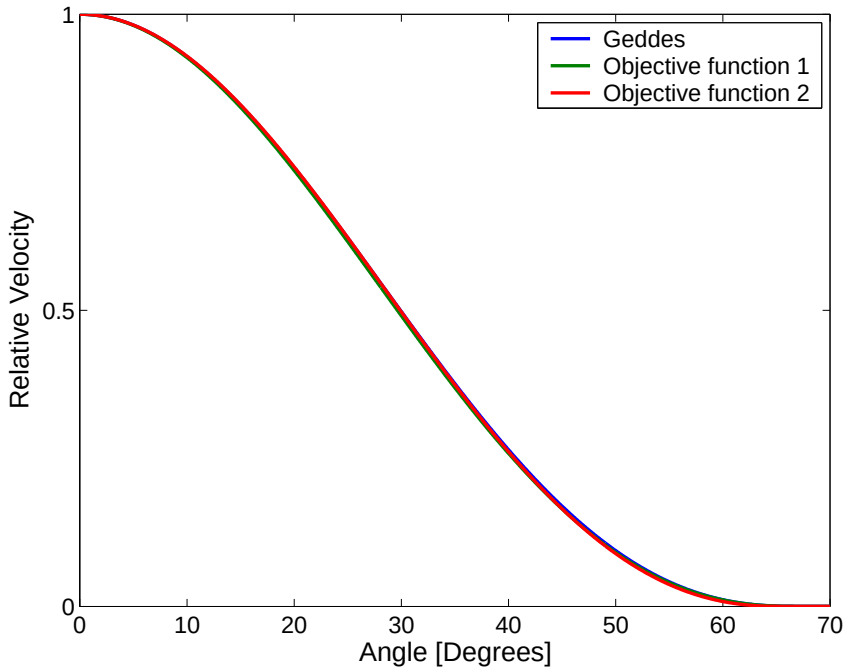
Table 6.3: Parameter values found using optimisation of different objective functions.

There is an increase in the number of function evaluations from about 250 (Table 6.2) to about 1500 (Table 6.3) when using exact rather than interpolated objective functions for objective function #1. This is possibly because this objective function has many more local minima than shown in Figure 6.12, and the interpolation function used was not able to capture the fine detail.

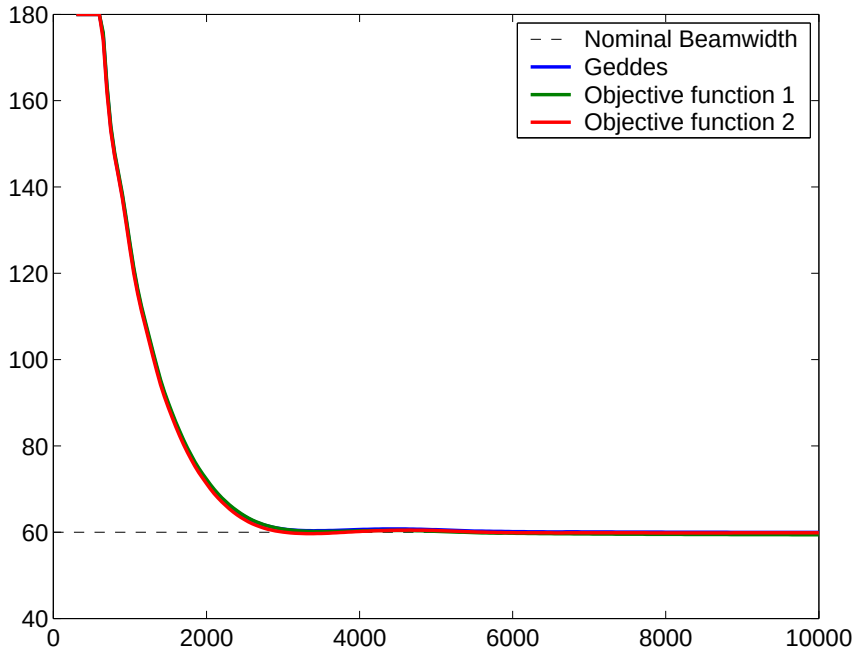
Figure 6.15 (a) shows the velocity profiles obtained by the optimisation techniques, com-

pared with the original Geddes profile, and Figure 6.15 (b) shows the beamwidths obtained. Both the profiles and beamwidths are virtually indistinguishable. Given the similarity of results obtained, and the difficulty that multiple local minima entail in finding a global optimum, and the efficiency of the equality constrained SQP algorithm, it is suggested that objective function #2 with equality constrained SQP be used for further optimisations.

In summary, two objective functions for finding an optimal beamwidth have been compared. Gradient based techniques (eg SQP) are not robust for objective function #1 and it requires a global optimisation techniques such as DIRECT. It is able to find a smooth beamwidth at a desired nominal value, however at the expense of an increase in the number of objective function evaluations. Objective function #2 with equality constrained SQP is able to find the best solution robustly for this simple problem. It is suggested that for more complicated problems that a hybrid technique be used, with the DIRECT global optimisation technique used for a few iterations to find a location close to the global minimum, and the SQP gradient based technique to refine the solution. The problem considered in this section is very simple, and more complicated problems with more variables must be examined before moving to optimisation of horn loaded loudspeakers. The use of SQP also requires the evaluation of the objective function gradient, as well as the objective function itself. The semi-analytical numerical method used in this section produces stable finite difference approximations to the gradient, as shown by the efficient performance of equality constrained SQP. Other numerical methods, such as the source superposition technique, may produce “noisy” numerical gradients (See Appendix B.2), hence the development of efficient gradient free optimisation techniques are required.



(a) Velocity profile



(b) Beamwidth

Figure 6.15: Comparison between original Geddes velocity profile and the results found by optimising 2 different objective functions.

6.8 EGO optimisation

Objective function #2 developed in Section 6.7 is able to successfully find the optimal solution for a relatively simple problem using a gradient based Sequential Quadratic Programming (SQP) optimisation routine. Whether this is able to be scaled to more complicated and larger dimensional problems, possibly with noisy gradient information, is another question. For the current application, the calculation times for each objective function are large, and there has been much research into optimisation of problems where the evaluation of the objective function is computationally expensive (Booker et al., 1998). The main thrust of many of these techniques is to use a surrogate, or meta-modelling technique (Appendix B.4) where an expensive objective function evaluation is replaced with a technique that interpolates between sparse samples with low computational overhead. In this thesis, the DACE surrogate modelling technique is adopted. DACE, or Design and Analysis of Computer Experiments, takes its name from a seminal paper by Sacks et al. (1989), and has become widely adopted. DACE uses a statistical interpolation technique called Kriging, and the two names are used interchangeably in this thesis. It is used for computer simulations where the results are deterministic, i.e. for the same inputs the simulation gives the same output, which is a different philosophy from that of Design of Experiments (Myers and Montgomery, 1995), where repetition of input conditions is used to reduce random error. The Kriging method is described in more detail in Appendix B.4 but for the purposes of this thesis, Kriging can be considered as an efficient interpolating technique that provides both a prediction of the mean value and the mean square error between the known sample results.

Since the Kriging interpolation technique is computationally inexpensive to evaluate, the question then becomes how best to sample the expensive objective functions, to allow efficient interpolation without sacrificing accuracy, and missing the global minimum. One approach is to sample the function on a rectangular grid. For optimisations involving dimensions greater than 2, the cost of directly sampling the entire objective function space

becomes prohibitively expensive. For example, to fit a Kriging model using a rectangular grid spacing of 25 points varied linearly between the lower and upper bounds for each dimension, the total number of points in two dimensions is $25^2 = 625$. Given the results found in Section 6.7, it would have been more prudent to not use any form of Kriging model and directly optimise the cost functions. However for the purposes of investigating different objective and constraint functions, the Kriging model allows rapid evaluation. The Kriging model was also used to overcome difficulties in applying the constraint function using the MATLAB SQP implementation, `fmincon`. In three dimensions, which would be required if there were three parameters to optimise, the number of samples jumps to $25^3 = 15625$, which is prohibitively expensive.

An alternative solution would be to try to sample the points more efficiently. The Latin Hypercube Sampling (LHS) strategy of McKay et al. (1979) is one such approach. It is a constrained Monte-Carlo sampling technique that improves the sampling efficiency over purely random (Monte-Carlo) sampling. Other techniques build on this approach, and the one used in this thesis is the space filling Improved Distributed Hypercube Sampling (IHS) of Beachkofski and Grandhi (2002). For a more complete description of LHS and IHS see Appendix B.5.

Whilst IHS will find a distribution of initial points that is more efficient at sampling higher dimensions than a hyper-rectangular grid, once sampled, it is still not known how accurately the Kriging meta-model represents the objective functions. Approaches such as cross validation (Schonlau, 1997, Section 4.3) and bootstrapping (den Hertog et al., 2004) go some way to ascertaining the accuracy of the Kriging model; however an automated approach would be helpful.

The EGO and SuperEGO algorithms of Schonlau (1997) and Sasena (2002) respectively go one step further than automatically improving the accuracy of a Kriging meta-model. They search for a global minimum by using information about the error in the Kriging approximation away from the sample points. They do this by performing a secondary

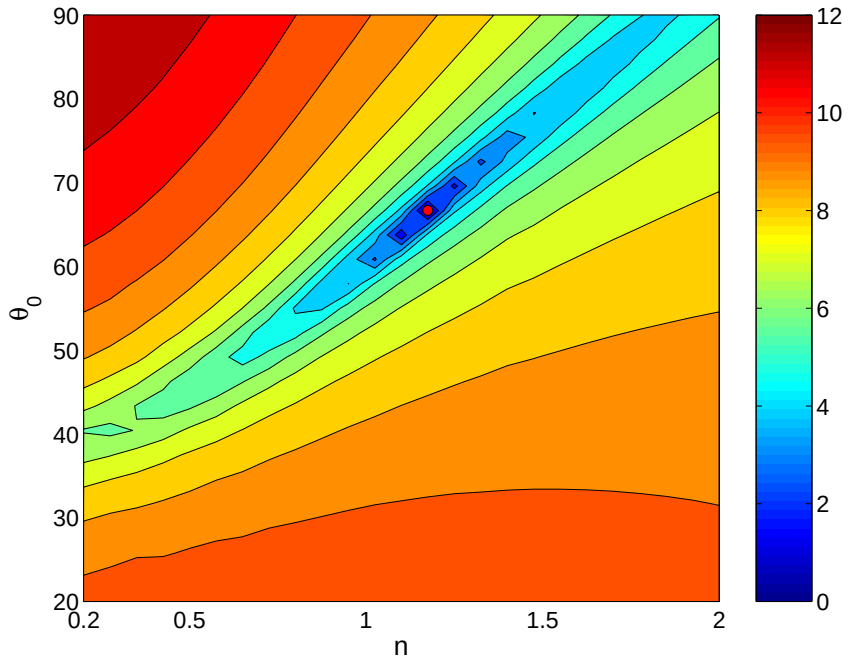
optimisation on an auxiliary problem, defined as “find the next best point”. This is called the Infill Selection Criterion (ISC) from the geostatistical literature and these techniques use a statistical approach to find the next point that is most likely to improve the current solution. Various forms of ISC are available with different emphasis on global and local searching. A simple example of EGO appears in Appendix B.6, along with further explanation.

6.8.1 Kriging interpolation of objective function values

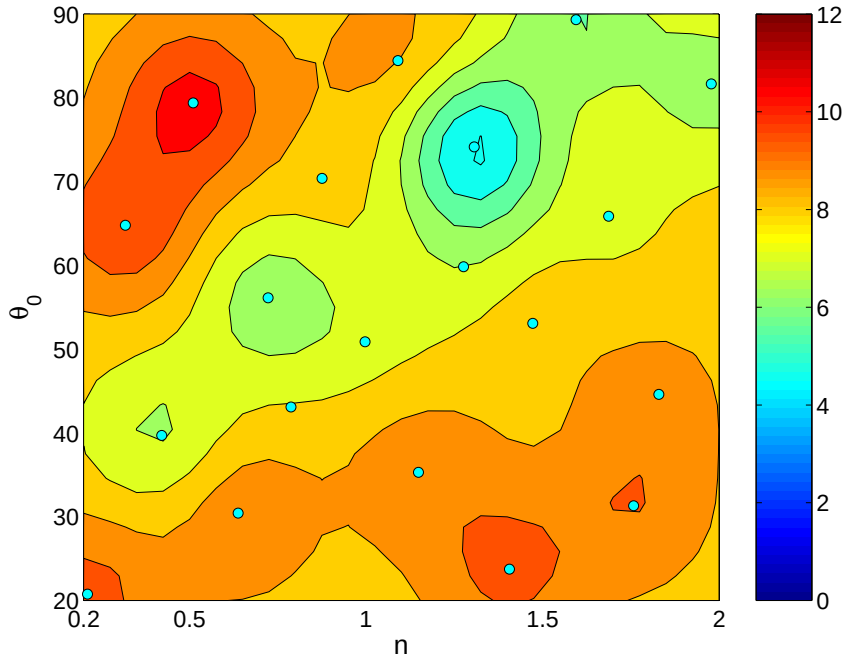
As an illustration of the process of the Kriging interpolation used in the EGO optimisation method, the objective function #1 of Section 6.9 is used as an example. The IHS sampling method was used to generate 20 initial points, and the function Φ_1 was calculated at these points. A Kriging model was calculated for this very sparse point spacing for each function, and both the full model and the approximation appear in Figure 6.16. The points sampled using the IHS method are shown as cyan dots. The Kriging approximation initially appears to be a poor approximation to the original function, however on further observation, it can be seen that the same general trends are evident.

Figure 6.17 shows contours of Φ_2 , the mean value of the beamwidth, for both the full function, and the sparse Kriging approximation. The sparse sample points are shown as cyan dots. Figure 6.18 shows similar contours for Φ_3, the standard deviation of the beamwidth. Both Φ_2 and Φ_2 are used in objective function #2. Again the sparsely sampled results show similar trends to the full approximation, giving hope that the EGO algorithm will be able to find an appropriate solution.

By design, the values of the Kriging surrogate model at the sample points are equal to the values at the sample points and Kriging is a true interpolation method. The advantage of EGO is that the Kriging approximation only has to be good enough to indicate where to sample next.

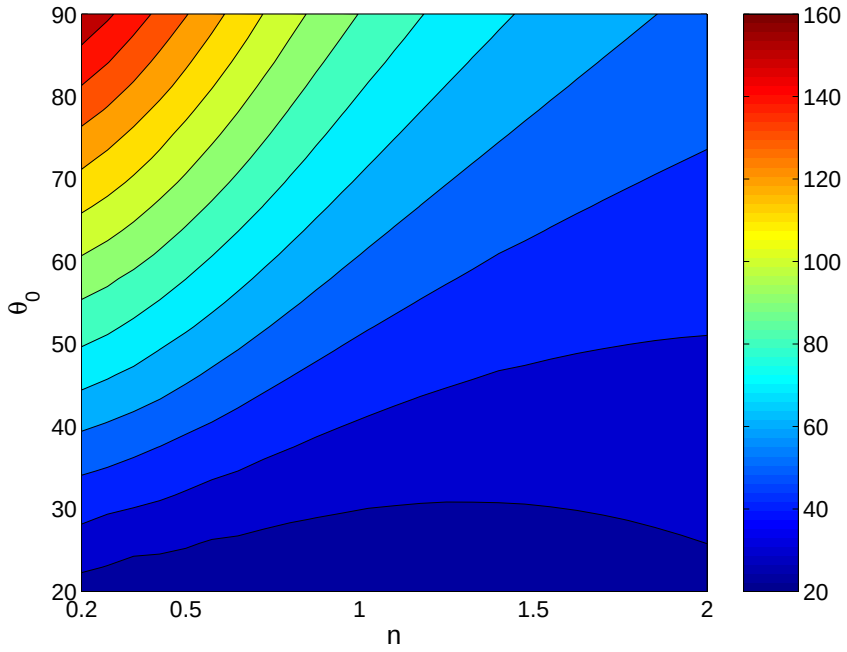


(a) Full function

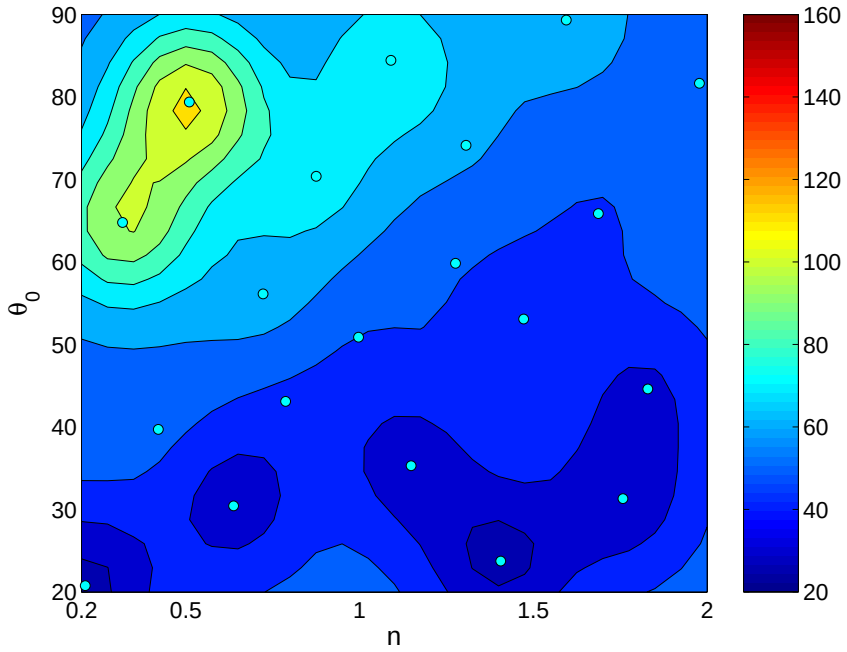


(b) Sparse Kriging approximation

Figure 6.16: Contours of $\ln\Phi_3$*, logarithmic least squared objective function. Cyan dots show the position of the initial sample points.*

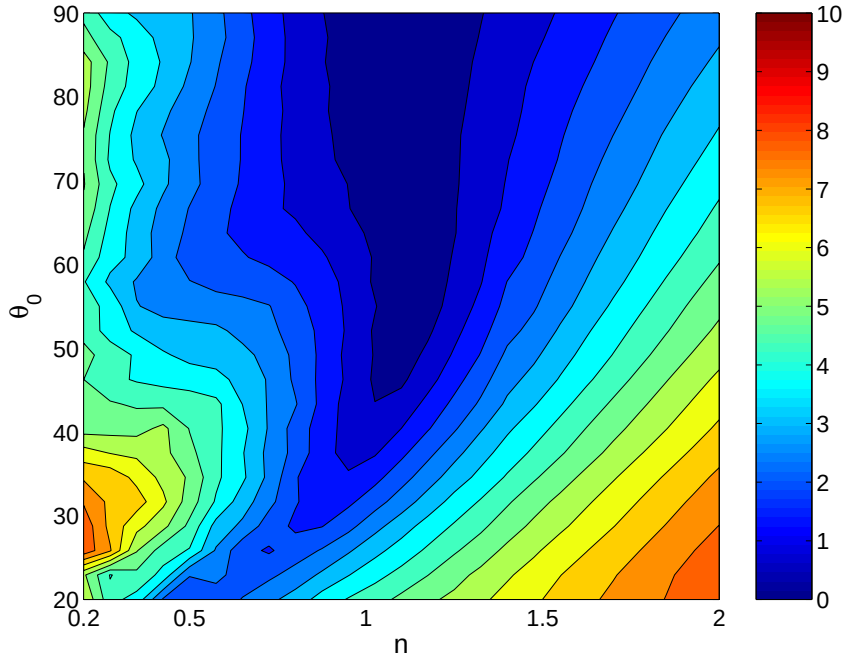


(a) Full function

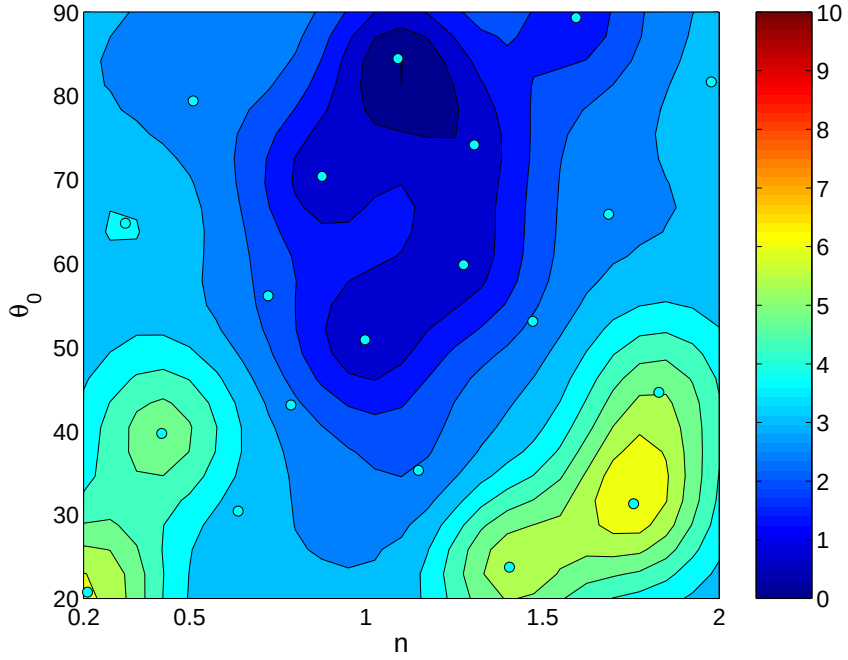


(b) Sparse Kriging approximation

Figure 6.17: Contours of Φ_2, mean value of the beamwidth.



(a) Full function



(b) Sparse Kriging approximation

Figure 6.18: Contours of Φ_3, standard deviation of the beamwidth.

6.8.2 Objective function #1

The application of the EGO algorithm with the Expected Improvement (EI) Infill Sampling Criterion (ISC) for 80 iterations (for a total of 100 function evaluations) to Objective function #1 is now examined.

The Expected Improvement (EI) function, given in Equation B.4 in Appendix B.6 and repeated here as Equation 6.16, is able to find the next sample point that is most likely to be an improvement over all previous samples. The EI ISC is defined as the expected value of the improvement in the current prediction at a point, over the minimum value of the current set of samples and is given as

$$EI = \mathbf{E}[I] = (f_{min} - \hat{y}(x))\,CDF_{\Phi_3}(f_{min}) + s(x)\,PDF_{\Phi_3}(f_{min}) \tag{6.16}$$

where $\hat{y}(x)$ and $s(x)$ are the mean and estimated standard deviation of the prediction of variable Φ_3 at point x, PDF_{Φ_3} is the probability density function of variable Φ_3 and CDF_{Φ_3} is the cumulative density function of variable Φ_3.

Unlike other ISC described in the literature (Jones, 2001), the Expected Improvement (EI) function is able to search implicitly for points likely to yield a local improvement where uncertainty is low, but the objective function is also near a minimum, while retaining the ability to search for global improvement where uncertainty in the sampling is high.

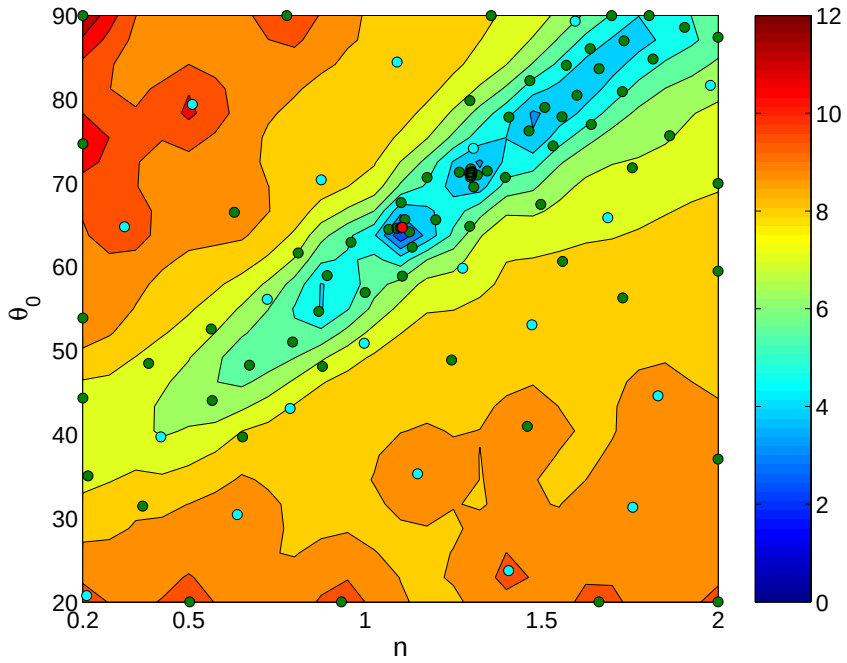
In Figure 6.19, the cyan dots show the position of the initial sample points, the green dots show the position of the samples chosen by the EGO algorithm using the EI ISC, and the red dot shows the location of the best sample. Figure 6.19 shows the results of two runs of the optimisation.

Table 6.4 gives the position of the best solutions found by the two runs of the EGO optimisation, and compares it to the values found by the DIRECT optimisation technique used in Section 6.7 for objective function #1. The values found by the EGO optimisation

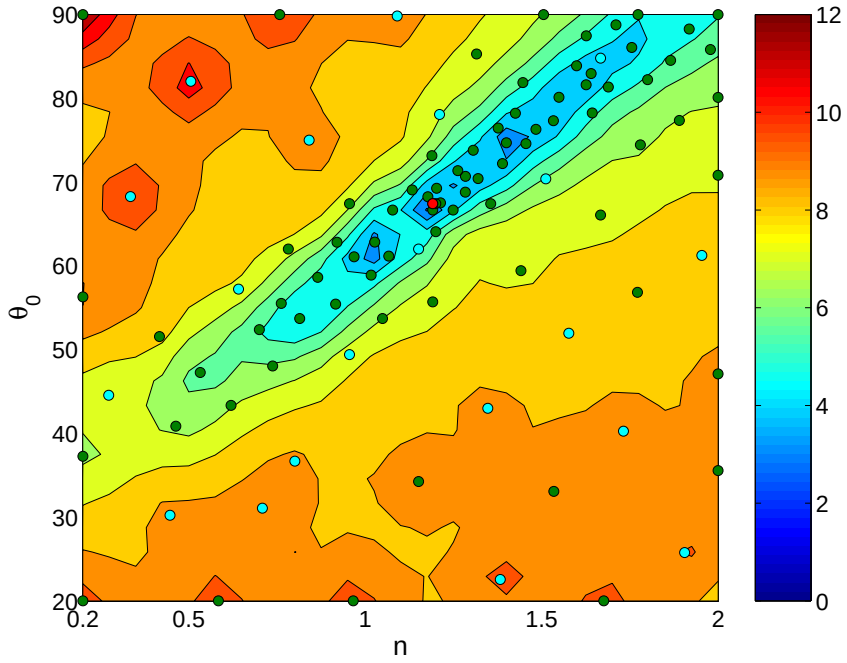
highlight the multiple local minima nature of the least squares objective function because each of the two EGO optimisation runs have reached different local minima. The DIRECT method gives the best solution, measured by the lowest value of $\ln(\Phi_1)$. The EGO technique is sensitive to the choice of initial samples, although the method will find the optimum given enough iterations (in the limit of dense sampling, see Jones 2001, Section 7).

	n	θ_0	$\ln(\Phi_1)$	Φ_2	Φ_3	N
DIRECT	1.12	65.1	0.75	60.0	0.22	1538
EGO run #1	1.11	64.7	0.84	60.28	0.24	100
EGO run #2	1.19	67.4	0.91	60.0	0.39	100

Table 6.4: Optimal parameter values for EGO optimisation of objective function #1 *using Expected Improvement (EI) Infill Sampling Criterion (ISC).*



(a) Run 1



(b) Run 2

Figure 6.19: Contours of $\ln \Phi_1$ *the logarithmic least squared objective function after application of EGO algorithm. The cyan dots show the position of the initial sample points, the green dots show the position of the samples chosen by the EGO algorithm using the EI ISC, and the red dot shows the location of the best sample.*

6.8.3 Objective function #2

Objective function #2, a constrained minimisation, was found to be very efficient when used to directly optimise the cost function. This section examines the behaviour of objective function #2 with EGO optimisation. There are a number of different methods of applying constraints to the EGO algorithm, including a probabilistic constraint method (Schonlau, 1997) and a constrained ISC subproblem (Sasena, 2002). These methods are examined in this section.

Probabilistic constrained Expected Improvement (EI) ISC

The method of Schonlau (1997) turns a constrained optimisation into an unconstrained optimisation by multiplying the expected improvement criterion by the probability of the constraint being active. The constraint used is Equation 6.14 and the probability of this constraint occurring if Φ_2 is a random variable is given by

$$P\left(\mathcal{B}_{nom}-\varepsilon \leq \Phi_2(x) \leq \mathcal{B}_{nom}+\varepsilon\right) = CDF_{\Phi_2}\left(\mathcal{B}_{nom}+\varepsilon\right) - CDF_{\Phi_2}\left(\mathcal{B}_{nom}-\varepsilon\right) \qquad (6.17)$$

where CDF_{Φ_2} is the cumulative density function of variable Φ_2.

Figure 6.20 shows the results of EGO inequality constrained optimisation of objective function #2 with $\mathcal{B}_{nom} = 60^\circ$ and $\varepsilon = 1^\circ$ using Schonlau's probabilistic constraint method. The expected improvement infill selection criterion is used, and the optimisation is run for 80 iterations, for a total of 100 objective function evaluations. Figure 6.20 shows initial samples with cyan dots, the ISC samples with green dots and the best sample with a red dot. The optimisation has performed poorly, with many function evaluations outside the constrained area, and no convergence to a final solution. The best solution obtained is reported in Table 6.6 with the label "Probabilistic constrained EI ISC".

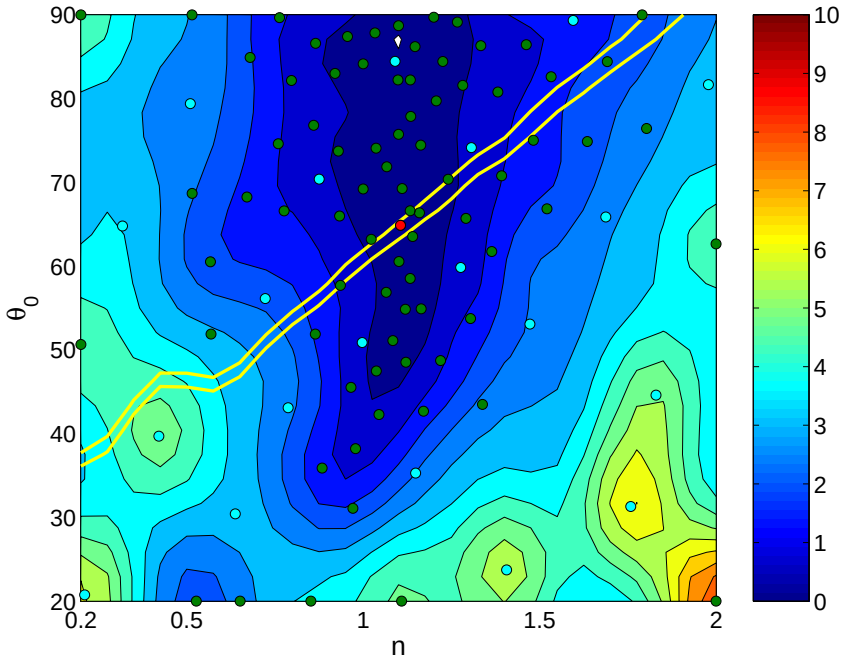


Figure 6.20: EGO optimisation of objective function #2 using probabilistic constraint method. Initial samples are shown with cyan dots, ISC samples with green dots and the best sample with a red dot. The yellow lines represent the constraint bounds.

Constrained Expected Improvement (EI) ISC

Sasena (2002) has developed a constrained EGO optimisation method by using a constrained ISC technique. This method applies constraints directly to the ISC subproblem, and uses the mean value of a Kriging approximation to the constraints when the evaluation of the constraint function is expensive. It has been found to be effective with the "regional extreme" criterion of Sasena. This criterion is not used in this thesis for reasons described in Appendix C. The expected improvement criterion was not found to perform well by Sasena with constrained optimisation using the DIRECT method, and this is confirmed in Figure 6.21. The initial samples are shown with cyan dots, the ISC samples with green dots and the best sample with a red dot. This minimum point is given in Table 6.6, "Constrained EI ISC". The points sampled by the ISC mainly fall within the constraints, but there is no convergence or repeated sampling of results near the minimum. The arrow in Figure 6.21 shows where repeated sampling has occurred, well away from the minimum.

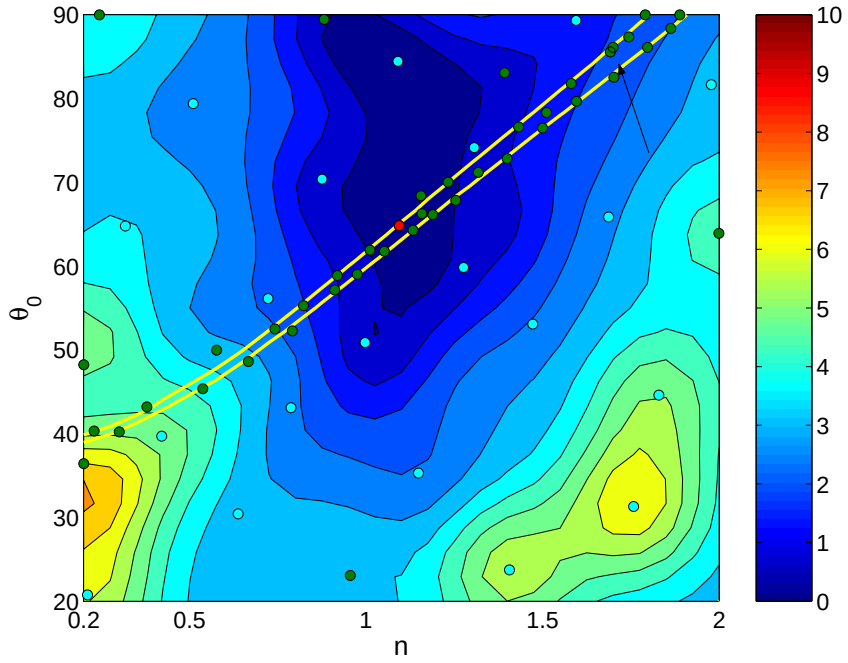


Figure 6.21: EGO optimisation of objective function #2 using ISC constraint method with expected improvement ISC. Initial samples are shown with cyan dots, ISC samples with green dots and the best sample with a red dot. The yellow lines represent the constraint bounds, and the arrow shows where repeated sampling occurs.

Both these methods of constraint appear to have found a value close to the minimum by essentially random sampling rather than some form of systematic search. Due to the poor performance of both the Probabilistic constrained Expected Improvement (EI) and the Constrained Expected Improvement (EI) Infill Sampling Criterion (ISC), alternative ISC were sought.

Constrained Minimum Objective Function (MOF) (ISC)

Figure 6.22 shows the results of 20 initial sample with a further 20 iterations of the ISC constrained EGO algorithm with a Minimum Objective Function (MOF) infill sampling criterion. This criterion does not try to do any kind of global searching and simply minimises the mean value of the Kriging predictor, i.e.

$$MOF = \hat{y}(x) \tag{6.18}$$

where $\hat{y}(x)$ is the mean value of the Kriging predictor.

For this example, the results are very impressive, with the minimum value being found in under 40 total objective function evaluations. The initial sample positions appear as cyan dots, the green dots represent the minimum objective function sampling and the red dot the optimum. This compares well with the performance of gradient based methods in the previous sections. The best result from the optimisation is given in Table 6.6 ,"Constrained MOF ISC". It is presumed that the results will be quite sensitive to the choice of initial sample points, and that there would be no chance of achieving a global minimum if the problem was not as simple as that posed in this section.

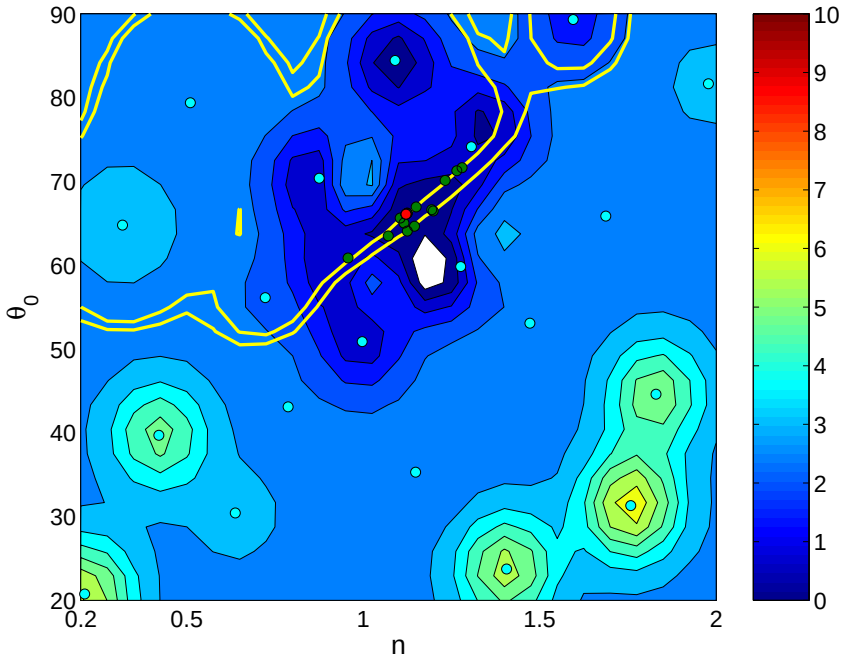


Figure 6.22: EGO optimisation of objective function #2 using ISC constraint method with minimum objective function ISC. The initial sample positions appear as cyan dots, the green dots represent the minimum objective function sampling and the red dot the optimum. The yellow lines represent the constraint bounds.

Constrained MAXimum VARiance (MAXVAR) / Minimum Objective Function (MOF) ISC

A solution to the potential lack of global optimum robustness in the Constrained Minimum Objective Function (MOF) ISC is to first focus the sampling to the feasible regions within the constraints, and then to perform an optimisation once the feasible region is sufficiently sampled. This is essentially a one step "switching" criterion as described by Sasena (2002). For the current problem, the global searching can be done efficiently because the constrained area is a small subset of the total search space. The optimisation proceeds in two stages. First, a constrained optimisation with an infill sampling criterion that finds the MAXimum VARiance (MAXVAR) efficiently samples the feasible space for 20 iterations. The MAXVAR criterion is

$$MAXVAR = -s(x) \tag{6.19}$$

where $s(x)$ is the estimated standard deviation of the prediction at point x. Minimising the negative of the standard deviation is the same as maximising the standard deviation.

Once the MAXVAR sampling is completed, a further optimisation using the Minimum Objective Function (MOF) efficiently finds the minimum within a further 20 iterations, for a total of 60 objective function evaluations. The results of this simulation appear in Figure 6.23 and in Table 6.6 ,"Constrained MAXVAR/MOF ISC". The initial sample positions appear as cyan dots, the points selected by the constrained maximum variance sampling phase appear as yellow dots, the green dots represent the minimum objective function sampling, and the red dot the optimum. Most of the final stage optimisation (the green dots) are tightly clustered around the optimal solution (the red dot), indicating a converged solution. The optimum result is the same as the minimum objective function criterion, but have added robustness because the entire feasible region is efficiently scanned.

For higher dimensional problems, the initial sampling may be very sparse. The perfor-

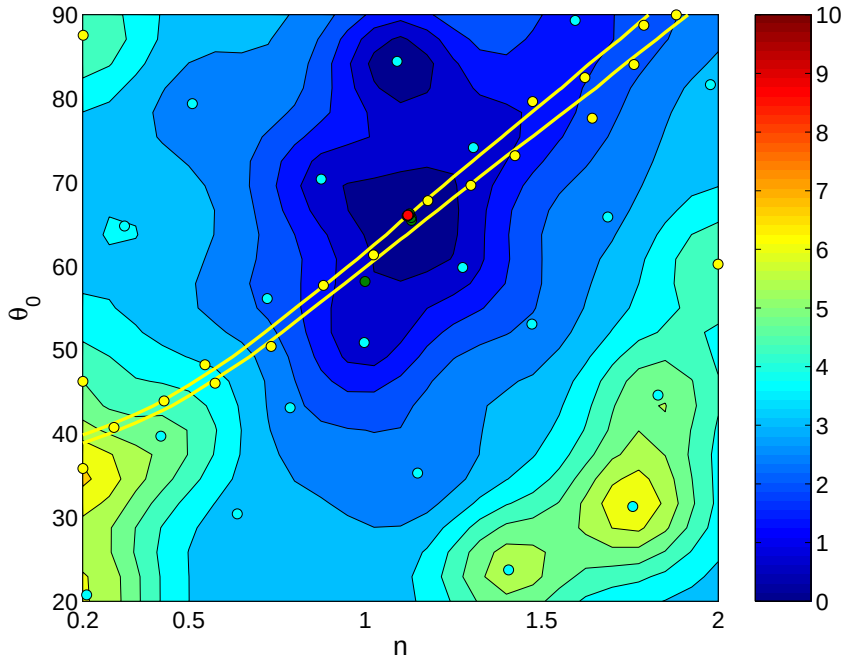


Figure 6.23: EGO optimisation of objective function #2 using ISC constraint method with MAXVAR ISC followed by MOF ISC. The initial sample positions appear as cyan dots, the points selected by the constrained maximum variance sampling phase appear as yellow dots, the green dots represent the minimum objective function sampling, and the red dot the optimum. The yellow lines represent the constraint bounds.

mance of the optimisation algorithm with a limited number of initial sample points (5) is examined in Figure 6.24. The colouring of the sample points is the same as in Figure 6.23, and the optimum result found is the same as previous optimisation ("Constrained MAXVAR/MOF ISC" in Table 6.6).

Table 6.6 summarises the results obtained by various methods of constrained optimisation of objective function #2 using the EGO method. Results obtained by the equality constrained SQP from Table 6.3 are also presented again for reference. It can be seen that the MAXVAR/MOF and MOF results have achieved the lowest value of Φ_3, and by this measure have the best performance. However, due to the nature of the inequality constraint used, the nominal beamwidth Φ_2 is not exactly 60°, and the measure $\ln(\Phi_1)$ is higher than that of equality constrained SQP. This issue can be resolved by decreasing the constraint tolerance parameter, ε, although if this is too small, the DIRECT method used to find the minimum of the Infill Sampling Criterion (ISC) will have convergence

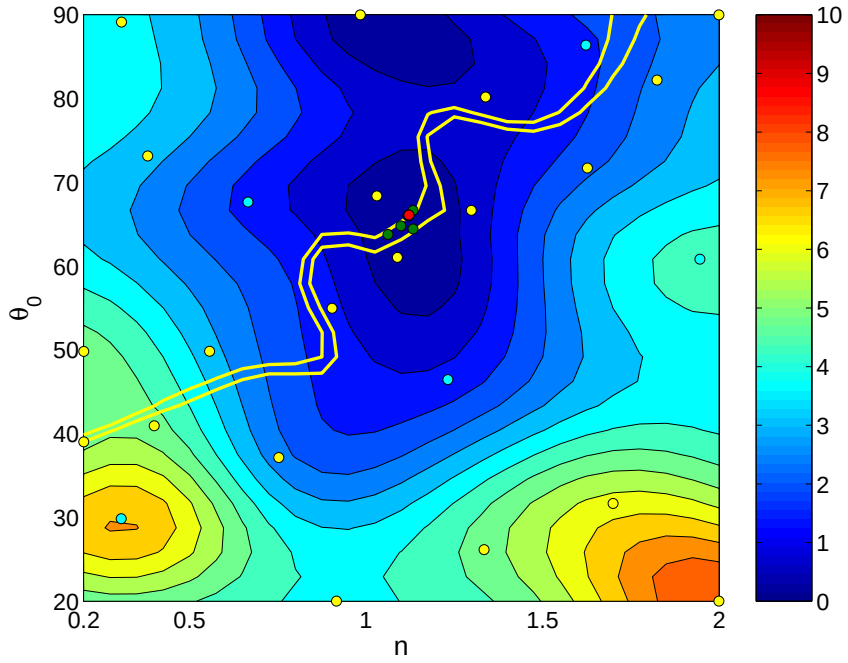


Figure 6.24: EGO optimisation of objective function #2 using ISC constraint method with maximum variance ISC followed by minimum objective function ISC. This optimisation has a limited number of initial sample points (5). The initial sample positions appear as cyan dots, the points selected by the constrained maximum variance sampling phase appear as yellow dots, the green dots represent the minimum objective function sampling, and the red dot the optimum. The yellow lines represent the constraint bounds.

problems. The degree of constraint compliance is probably not an issue when comparing the accuracy of the model with experimental measurements of beamwidth.

	n	θ_0	$\ln(\Phi_1)$	Φ_2	Φ_3	N
Equality constrained SQP	1.13	65.2	0.83	60.0	0.22	61
Probabilistic constrained EI ISC	1.11	64.9	1.3	60.4	0.23	100
Constrained EI ISC	1.09	64.8	2.35	60.7	0.26	100
Constrained MOF ISC	1.12	66.1	2.87	61.0	0.21	40
Constrained MAXVAR/MOF ISC	1.12	66.1	2.87	61.0	0.21	60

Table 6.5: Optimal parameter values for constrained EGO optimisation of objective function #2 using different Infill Sampling Criterion (ISC).

In summary, a number of techniques that can efficiently and robustly find the global optimum of expensive objective functions have been investigated. It has been found EGO optimisation of objective function #2, the constrained minimisation of the "smoothness" of the beamwidth is the most efficient. When a constrained MOF ISC is used, the technique

can be made robust by additional sampling of the feasible region using a MAXVAR ISC. However the amount of additional sampling required for robustness is not known *a priori* and is quite problem dependent. Further research is required to make the constrained ISC subproblem a robust general global optimisation technique. This current technique for the problem considered in this thesis should be investigated further with more difficult optimisation problems.

6.9 Optimised Bézier velocity profile

A velocity profile described by a Beézier curve (See Appendix D) gives a much greater control over the shape than Equation 6.8. This control is seen in Figure 6.25 with an independent, two parameter, control of the upper profile curvature and the lower tail. Each profile has been normalised to a 30° nominal design angle, the angle at which the velocity profile reaches $1/2$ of the on-axis value.

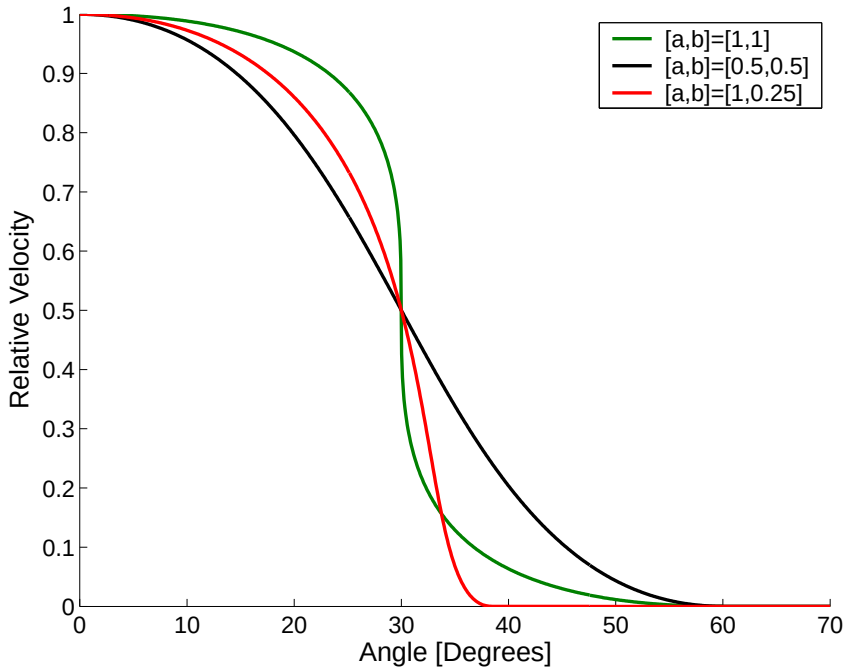
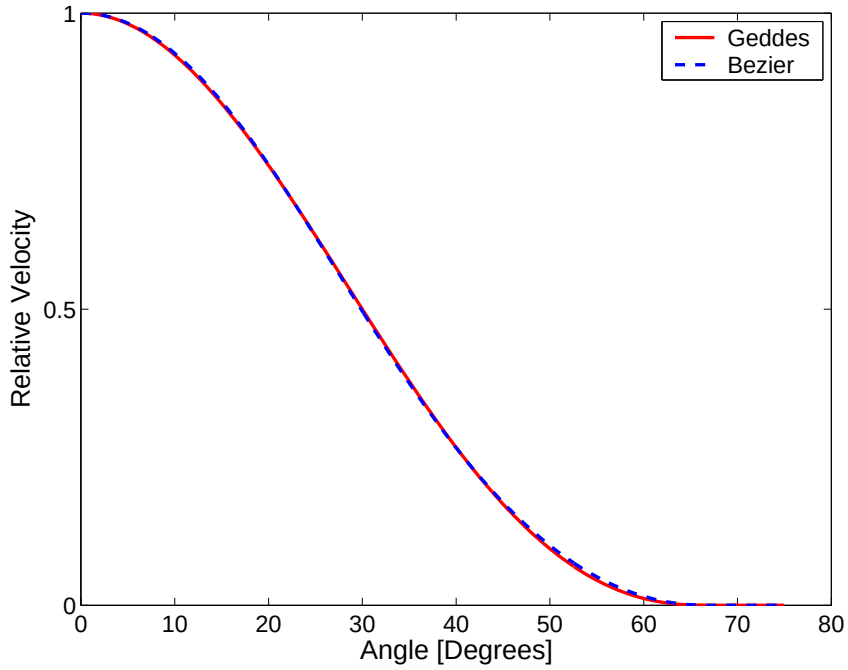


Figure 6.25: Comparison of velocity profiles for different parameters of the Bézier profile.

To test the optimisation method, the parameterisation of the geometry profile must be able to reproduce an optimal, or near optimal, profile such as that of Geddes (Equation 6.6). These optimal parameters, $\{a, b, \theta_0\} = \{0.33, 0.47, 66.5\}$, found using a numerical fitting procedure, make the Bézier profile as close as possible to the Geddes profile, as shown in Figure 6.26.

Because the Bézier profile has more freedom to define the velocity profile, it is a more difficult problem to optimise. Conversely, as it has more freedom it has more potential to find a better solution. For problems where the gradient can be calculated, Section 6.7



(a) Beamwidth

Figure 6.26: Comparison of Bézier and Geddes profiles, showing that the Bézier profile has the ability to approximate the Geddes profile very well.

found that the Sequential Quadratic Programming (SQP) algorithm with equality constrained objective function #2, after an initial search with the DIRECT algorithm is an efficient method. For the 3 parameter Bézier profile, equality constrained SQP produces an optimal solution in $N = 120$ objective function evaluations, with results given in Table 6.6 with the label "SQP optimum". The original Geddes profile numbers are included for reference.

	a	b	θ_0	$\ln(\Phi_1)$	Φ_2	Φ_3	N
Geddes	0.33	0.47	66.5	1.65	60.1	0.57	n/a
SQP optimum	0.36	0.46	65.2	0.56	60.0	0.19	120
EGO optimum	0.36	0.45	65.6	1.63	60.5	0.18	100

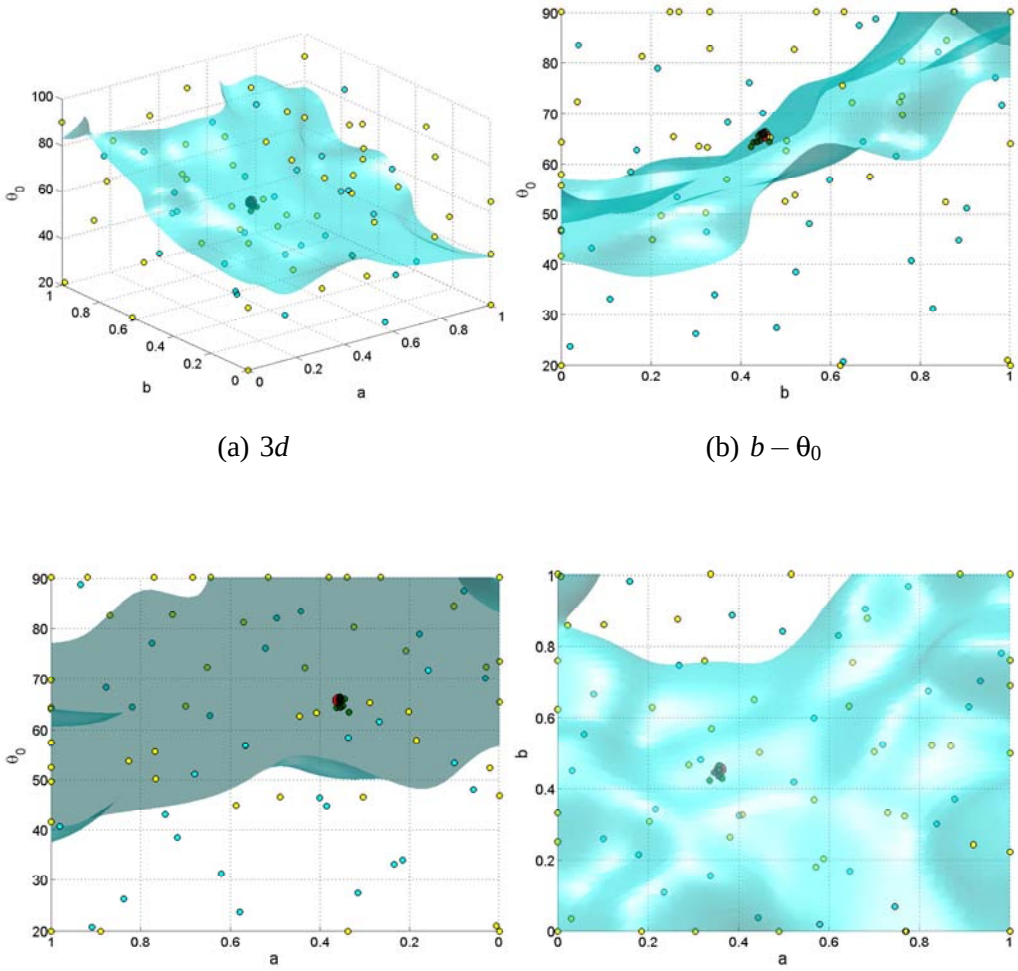
Table 6.6: Optimal parameter values for Bézier profile optimisation.

For optimisations when the gradient is not easily calculated, the Enhanced Global Optimisation (EGO) technique with objective function #2, described in Section 6.8, can be

used. A slight modification of the algorithm uses Adaptive Simulated Annealing (ASA) (Ingber, 1993) as the optimiser for the ISC. This method was found to be more robust than the DIRECT method for this problem.

A visualisation of the sample points chosen by the optimisation technique after 100 objective function calculations is given in Figure 6.27 in 3 dimensions. The calculation of 30 initial samples, shown by the cyan dots, was followed by the calculation of 50 samples selected by the constrained MAXVAR (yellow dots), and then 20 samples using MOF (green dots). The optimum is shown by the red dot, with repeated sampling near the optimum indicating some sort of convergence. A surface of constant $\Phi_1 = 60^\circ$ appears in blue. The optimum results appear in Table 6.6 with the label "EGO optimum".

Examination of Table 6.6 shows that both the SQP and EGO optimisations are able to find a Bézier parameterisation of the velocity profile with marginally better performance than the Geddes parameterisation. This is most probably due to the increased variability in the velocity profile allowed by the Bézier curves. More importantly, this section has shown that both the SQP and EGO optimisation techniques work on larger problems with more that 2 dimensions.



(a) 3*d* (b) $b - \theta_0$

(c) $a - \theta_0$ (d) $a - b$

Figure 6.27: Sampling of parameter space during EGO optimisation.

6.10 Conclusions

A “Constant Beamwidth Transducer (CBT)” has been introduced, which is able to produce, with a special velocity profile over the surface of a sphere, a frequency independent beamwidth with a smooth frequency response. These are desirable characteristics for a horn used to produce sound in cinemas. The theory and literature on the CBT and also on other methods of producing frequency independent beamwidths, including the velocity profile of Geddes (2002) was reviewed.

Next, a semi-analytical calculation technique to calculate the beamwidth for a given velocity profile was developed. This was used to calculate four different velocity profiles: a constant velocity over a spherical cap mounted on the surface of a sphere; a CBT profile; the profile of Jarzynski and Trott; and the Geddes velocity profile. This simulation showed that the Geddes velocity profile produces the smoothest beamwidth response, possibly at the expense of low frequency performance. The performance of the different velocity profiles was examined in the context of CBT theory, with the best performing profiles showing the highest rate of energy decay in the spherical Legendre modes.

CBT theory also suggests that the performance of CBT transducers is unaffected by the removal of the inactive part of the sphere, i.e. the part over which the velocity profile is zero. This is confirmed numerically by simulations using the source superposition technique.

Finally, the optimisation of the velocity profile is considered. Two different objective functions are described, one that uses least squares to drive the velocity profile to find the best beamwdith performance, and the other that uses a constrained optimisation of a smoothness parameter. For simulations where gradient information is readily available, it was found that equality constrained Sequential Quadratic Programming (SQP), with an initial search using the DIRECT global optimisation method performed best, and was able to find an optimal solution in an acceptable number of objective function evaluations.

For simulations where gradient information is unavailable or "noisy", the Enhanced Global Optimisation technique was able to find an optimal solution in an acceptable number of objective function evaluations. It does this by sampling the parameter space using a space filling method, then fitting a Kriging meta-model to describe both a prediction of the mean and the error of the objective function between the samples. An auxiliary optimisation is then performed using this efficient approximation to the true objective function to find the next best point to sample. A number of different auxiliary objective functions, called the Infill Sampling Criterion (ISC), have been tested with the two different primary objective functions. For constrained optimisation it was found that a strategy that firstly reduces the maximum error around the constraint, the MAXimum VARiance (MAXVAR) ISC, followed by minimising the mean value of the predictor (Minimum Objective Function or MOF) is an efficient method. For unconstrained problems, it was found that the Expected Improvement (EI) algorithm gave a good balance between local and global searching.

Two different parameterisations of the velocity profile were investigated. One parameterisation was similar to the Geddes velocity profile was described by 2 parameters, and the other, which allowed a more variable velocity profile, was described by Bézier curves and contained 3 parameters. Both equality constrained SQP and the EGO optimisation methods were able to find solutions that were better than the existing Geddes profile, although the improvements were marginal. More importantly, it shows that robust optimisation techniques are able to find global minima of expensive objective functions in a relatively small number of evaluations with larger dimensional problems.

The idea behind constant beamwidth transducers, as well as the robust optimisation techniques that have been developed, are now able to be applied to the optimisation of the profiles of horn loaded loudspeakers.

Chapter 7

Horn geometry optimisation

The work described in this chapter draws together the work of previous chapters to develop a method to optimise the geometry of a horn to give an easily specified frequency independent beamwidth and to provide a smooth frequency response over a large bandwidth, provided it is physically possible to do so.

The geometry of the horn is parameterised, and the source superposition technique that was verified in Chapter 4 and validated in Chapter 5 is used to calculate the beamwidth. Robust optimisation techniques, developed in Chapter 6 are then used to systematically modify the geometric parameters to find the optimum horn geometry.

7.1 Introduction

The Constant Beamwidth Transducer (CBT) investigated in Chapter 6 gives a frequency independent beamwidth. The sound field is produced by a specific, in phase, velocity profile defined over a small part of the surface of a sphere. Section 6.6 shows that the surface of the sphere over which the velocity is not defined is irrelevant to the performance of the transducer. This implies that if the mouth of a horn can generate the same velocity

field as that of the CBT over the surface of an imaginary spherical cap then the same, frequency independent, sound field will result.

Examination of the typical velocity profile required by a CBT, such as that shown in Figure 6.3, shows a gradual shading of the velocity profile away from the axis. Geddes (2002, Section 6.6) suggests that by flaring the exit of the horn, a change in the velocity profile, and hence beamwidth, will result. The tools to modify the flare of the horn in a systematic way have been developed, using the techniques outlined in previous chapters. Hence, the aim of the work described in this chapter is to use robust optimisation techniques to produce an optimal horn geometry that achieves constant beamwidth performance. It is also intended to investigate whether constant beamwidth performance can be achieved at a desired nominal beamwidth. To achieve this aim, the source superposition technique, which has been verified and validated in Chapters 4 and 5 respectively and is able to reproduce the sound field of a horn loaded loudspeaker, is coupled with the EGO optimisation technique, developed in Chapter 6, to find the smoothest design with constant beamwidth performance.

The original aims of this thesis were to provide a horn design procedure characterised by: an easily specified frequency independent beamwidth; and a smooth frequency response over a large bandwidth. If constant beamwidth performance is achieved, then CBT theory suggests that the second aim is automatically satisfied if the first is achieved, hence the focus on constant beamwidth behaviour in this thesis. This hypothesis will have to be tested either numerically or experimentally once a constant beamwidth horn design is achieved.

This chapter begins with an investigation of a geometrically simple horn profile consisting essentially of a conical horn, with a radiused entry at the horn throat and a radiused flare at the horn mouth. The ability of this geometry to achieve the desired nominal beamwidth is investigated, as is the the effect of throat radius on the performance of the system.

More complicated geometry parameterisations are then investigated, and a Bézier spline

based geometry is found to be flexible enough to find a shape that produces approximately constant beamwidth behaviour, although it may not be able to find a desired nominal beamwidth. This geometry parameterisation is then solved repeatedly for a wide range of lengths and throat dimensions, and the results used to develop a method that is able to quickly find an optimum horn design.

7.2 Optimisation method

The optimisation techniques developed in Chapter 6 can be used with the source superposition technique to calculate the optimum geometry profile for a given nominal beamwidth. It was found that objective function #2 provided a robust way of finding the optimum using constrained minimisation. This objective function is adopted here for the current problem, and is repeated here for clarity.

Objective function #2 (Equation 6.12) can be written as,

$$\min \Phi_3 \tag{7.1}$$

with equality constraint (Equation 6.13)

$$\Phi_2 = \mathcal{B}_{nom} \tag{7.2}$$

where Φ_2 and Φ_3 are defined as (Equations 7.3 and 7.4),

$$\Phi_2 = \texttt{mean}(\mathcal{B}(\mathbf{f} \geq f_{min})) \tag{7.3}$$

$$\Phi_3 = \texttt{std}(\mathcal{B}(\mathbf{f} \geq f_{min})) \tag{7.4}$$

where $\texttt{mean}(\mathbf{x})$ and $\texttt{std}(\mathbf{x})$ are the mean and standard deviation of vector $\mathbf{x}$, respectively, $\mathcal{B}(\mathbf{f})$ is a vector of beamwidths calculated using the source superposition technique over

a range of frequencies described by the vector **f**. The operator $\mathbf{f} > f_{min}$ selects only those frequencies above f_{min} and $\mathcal{B}_{nom}$ is the nominal (or desired) beamwidth.

In some cases, it may be necessary to relax the equality constraint to a bounded or pseudo-equality constraint such as,

$$\mathcal{B}_{nom} - \varepsilon \leq \Phi_2 \leq \mathcal{B}_{nom} + \varepsilon \tag{7.5}$$

where ε is a tolerance parameter that describes how close to the nominal beamwidth the optimisation is constrained.

Equation 7.5 can be reformulated as two inequality constraints,

$$\begin{aligned} \mathcal{B}_{nom} - \Phi_2 &\leq \varepsilon \\ \Phi_2 - \mathcal{B}_{nom} &\leq \varepsilon \end{aligned} \tag{7.6}$$

The objective function #2 (Equation 7.1) requires that the mean (Φ_2) and standard deviation (Φ_3) of the beamwidth be calculated over a range of frequencies, in this case from 3100 to 10000 Hz in steps of 400Hz, so $f_{min} = 3100$.

The definition of what constant beamwidth behaviour consists of is, however, not well defined. In this chapter, a horn shows constant beamwidth behaviour if the beamwidth approaches a constant nominal value smoothly from above (see for example the beamwidths shown in Figure 6.15 (b), produced by the Geddes velocity profile over the surface of a sphere). A more complete definition of what constitutes constant beamwidth behaviour should be investigated, and is recommended for future work (Section 8.3) in the context of defining more relevant objective functions.

It was suggested in Chapter 6 that a useful optimisation strategy would involve the use of the minimisation technique called Sequential Quadratic Programming (SQP). It was found that this technique was not suitable for the current problem because of difficulties

in calculating the gradient using finite differences (See B.2). The source superposition technique requires discretisation of a given geometry in order to calculate the beamwidth, and small changes in the mesh distribution or density can give relatively large changes in the beamwidth. This would be the case if a small change in curvature led to a change in the length of a line, which led to a step change in the number of elements used in the mesh. A small change of input can lead to a large change in output. The perturbation in parameters required by the finite different method was very sensitive to these changes and was not able to provide a sensible approximation to the gradient of the solution, hence a gradient free method such as DIRECT or EGO is required. Further research on an efficient method to calculate the gradient of the source superposition solution would be very valuable, and should be considered for future work (Section 8.3).

It was found in Chapter 6 that the DIRECT technique is not suitable for a constrained minimisation problem. It is, however, probably quite a good choice for an unconstrained minimisation of functions with a small number of local minima. It was found in Section 6.8.2 that the method required a large number of iterations with the many local minima of objective function #1.

The EGO optimisation technique (Chapter 6) is the best choice for constrained minimisation of the horn geometry problem because it was able to robustly find a constrained global minimum of objective function #2 with an acceptable number of objective function calculations, without requiring gradient information. A brief outline of the technique used here follows. The parameter space is sampled using Improved Hypercube Sampling (Beachkofski and Grandhi, 2002), the objective function evaluated at each sample point and then a Kriging method used to interpolate the data, as well as give an approximation to the error in the prediction. This information is then used to calculate the next best place to sample, by solving the Infill Sample Criteria (ISC) auxiliary problem. The interpolated surface is updated using the objective function calculated at the sample point, and the sampling process continued until a stopping criterion is reached (usually a maximum

number of samples).

The original EGO algorithm of Schonlau (1997) used an Expected Improvement ISC; however, this was found to be difficult to solve for a constrained minimisation (Sasena, 2002). A simple approach is adopted here. A number of iterations of solving a MAXimum VARiance (MAXVAR) ISC, with constrained minimisation, efficiently samples the feasible region of the objective function. This process aims to reduce the uncertainty in the interpolation of the objective function, but only within the constraint boundaries. A Minimum Objective Function (MOF) constrained minimisation samples the mean value of the interpolated objective function and then searches for a minimum value of the objective function within the constrained region. This technique can find a global minimum, provided that the sampling along the constraint is sufficient. The amount of sampling, which is problem dependent and cannot be determined a priori, is set by trial and error.

In Section 6.9 both the DIRECT technique and Adaptive Simulated Annealing (ASA) were used to solve the constrained ISC auxiliary problem. Both these techniques can require many thousands of function evaluations, and hence are not suitable for solving the objective function directly. When solving the ISC subproblem, the objective function is efficiently evaluated using Kriging interpolation, and the cost is minimal. For the problem described in Section 6.9, the DIRECT technique was found unsuitable and ASA was used. However, for general problems, neither technique has been found to be superior, and here a pragmatic approach is taken. Both the DIRECT technique and ASA are used to find a minimum of the ISC, using an inequality constrained approach. Then equality constrained SQP is used, starting with the solution of each global minimisation technique, to both enforce the equality constraint and improve the solution if possible. The best result is then used find the next sample position. There are no problems calculating the gradient numerically of the ISC using SQP[1]. This may seem complex, but the overhead required is

[1] It should be possible to derive an analytical gradient for all ISC objective functions as the MATLAB DACE toolbox predictor function supplies the gradient for both the mean prediction and the mean square error. It was found that numerical gradient estimations were efficient enough for the current problem.

minimal for small dimensions compared to the cost of calculating the objective function and it was found to provide a very robust search restricted to the feasible region.

7.3 Simple flared horn

A very simple horn profile is considered in this section. For comparison with the available existing small horns, the radius of the mouth of the horn is held constant (R_m = 82.5 mm) for all the simulations performed in this chapter. Keeping the mouth radius constant simplifies the construction of the cost function, and does not limit the analysis in any way as results can simply be scaled for different frequencies. The performance of horns with alternate mouth dimensions can be readily scaled from the existing results. Figure 7.1 shows a schematic of the geometry parameterisation. The angle of the conical horn, θ, and the length of the flange (F) govern the overall length of the horn as the mouth dimension is constant. A simple fillet of constant radius, that is tangent to both the conical horn and the end of the flange, flares the conical horn and provides some control over the horn mouth velocity profile. The radius of this curve is a function of the flange length. Finally, the horn is attached to the compression driver of radius R_t using another constant radius fillet.

7.3.1 Two inch throat

An optimisation of the geometry was undertaken for a constant horn throat diameter, $D_t = 2R_t$ = 50 mm (2 inches), a standard dimension for compression drivers. The desired nominal beamwidth was $\mathcal{B}_{nom} = 60^\circ$. The two free variables, θ and F (expressed as a percentage of the mouth radius), were normalised to lie between 0 and 1 using the upper and lower bounds given in Table 7.1.

Figure 7.2 shows the variation in geometry achieved by the parameterisation used. The

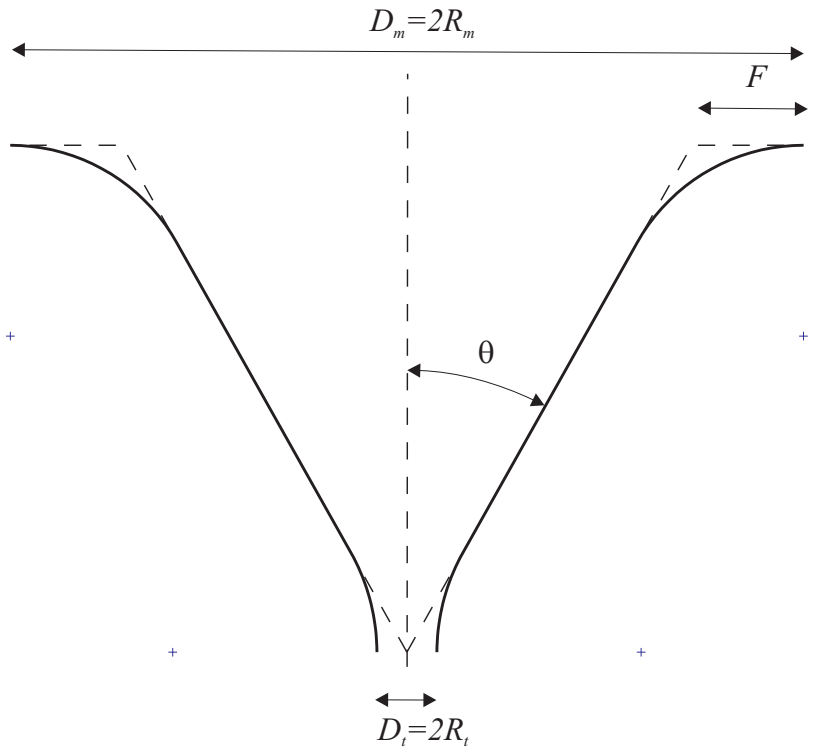


Figure 7.1: Simple horn geometry parameterisation. Consisting of a conical horn of angle θ *with a flange of length* F *with fillets of constant radius.*

Name	Variable	Lower	Upper
$x(1)$	θ	20°	40°
$x(2)$	$\frac{F}{R_m}$	5%	40%

Table 7.1: Upper and lower bounds of parameters used to describe the simple flared horn geometry.

parameters ($x(1)$ and $x(2)$, representing θ and F) are both systematically varied between the lower limit ($x=0$), the mid range value ($x=0.5$) and the upper limit ($x=1$), and the resulting horn profile shown.

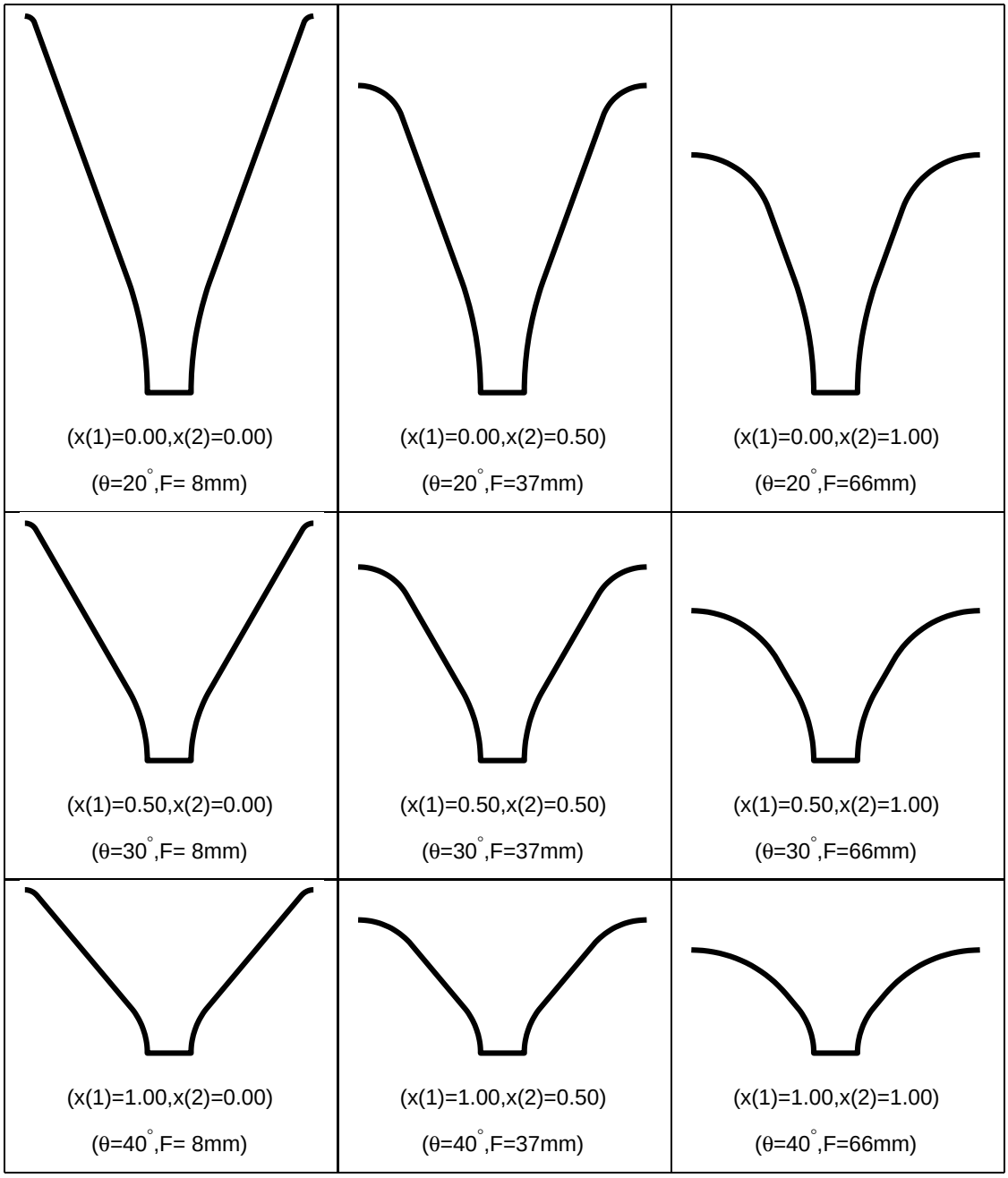


Figure 7.2: Variation in simple horn geometry with a 2 inch throat. Parameters vary between upper and lower bounds, $0 \leq x(1) \leq 1$ *and* $0 \leq x(2) \leq 1$.

Figure 7.3 shows the optimisation trajectory of an EGO optimisation. The filled contour is Φ_3, and the black line a contour of the constraint, $\Phi_2 = 60°$. The cyan dots are the positions of the initial samples (50) that initially characterise the objective function space. The yellow dots (50) show the MAXVAR sampling, used to reduce the uncertainty in

predictions, but only when the mean beamwidth constraint is achieved. The minimum objective function method is then used, and shown by the green dots (50). This finds the minimum, given by the red dot. It was found that the method repeatedly sampled the same position, which caused problems with robust fitting of the Kriging model (Lophaven et al., 2002b, Eldred et al., 2003). The solution to this problem is to merge the data set based on sample proximity before fitting the Kriging model. The MATLAB Dace toolbox function `dsmerge` was used with a coarse tolerance (1×10^{-2}).

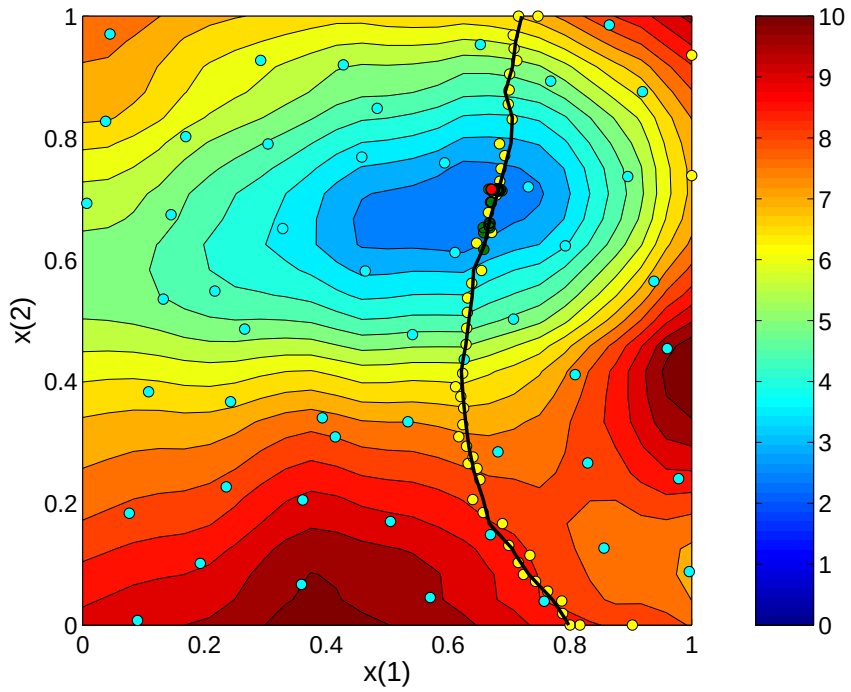


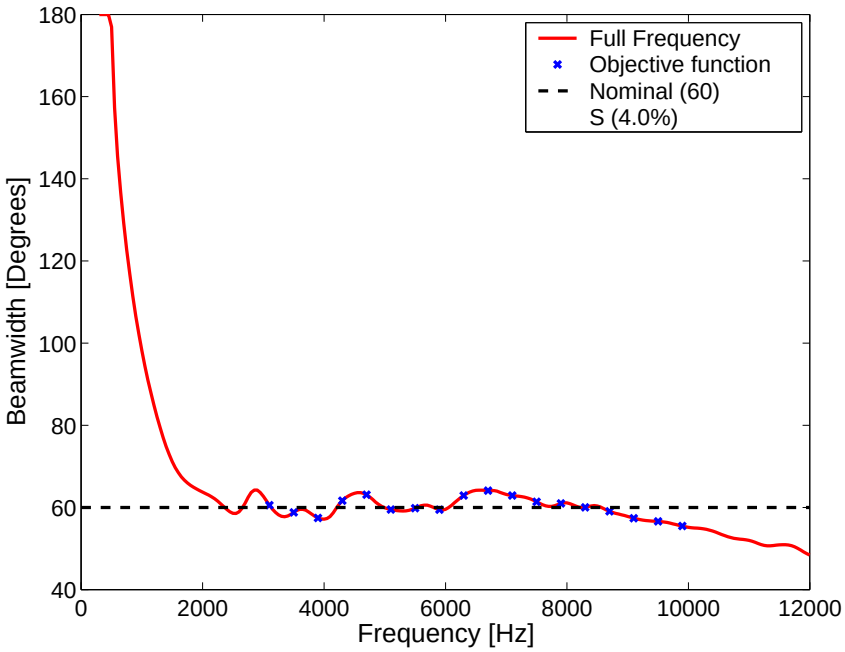
Figure 7.3: Optimisation trajectory for the simple horn geometry with a 2 inch throat. The cyan dots show the initial samples, yellow dots show the MAXVAR sampling, green dots the MOF sampling and the red dot shows the global minimum. The black line shows the constraint.

The numerical values of the optimal solution are $x(1) = 0.67$ ($\theta = 33^\circ$) and $x(2) = 0.72$ ($F = 50$ mm). The mean value of the beamwidth $\Phi_2 = 60^\circ$. The parameter S, defined as

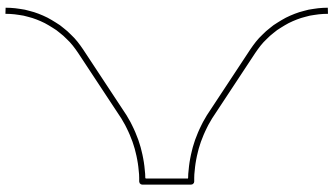
$$S = \frac{\Phi_3}{\Phi_2} \tag{7.7}$$

is an objective measure of how smooth the function is, where the smaller the value of S, the smoother the beamwidth over the range of frequencies considered. For the current

optimal solution, the value of S is 4%. The beamwidth produced by the optimisation is shown in Figure 7.4 (a) where the red line is the beamwidth evaluated over a full frequency range from 300 to 12000 Hz, the blue crosses are the frequencies over which the objective function is calculated, and the dashed line is the nominal beamwidth achieved (in this case 60°). The horn profile that generates this beamwidth is shown in Figure 7.4 (b). Even with this very simple geometry parameterisation, a reasonable approximation to constant beamwidth behaviour is achieved, if only over a limited bandwidth. The high frequency performance of this design decays with frequency above 9000 Hz, and this shows that true CBT behaviour has not been achieved.



(a) Beamwidth



(x(1)=0.67,x(2)=0.72)

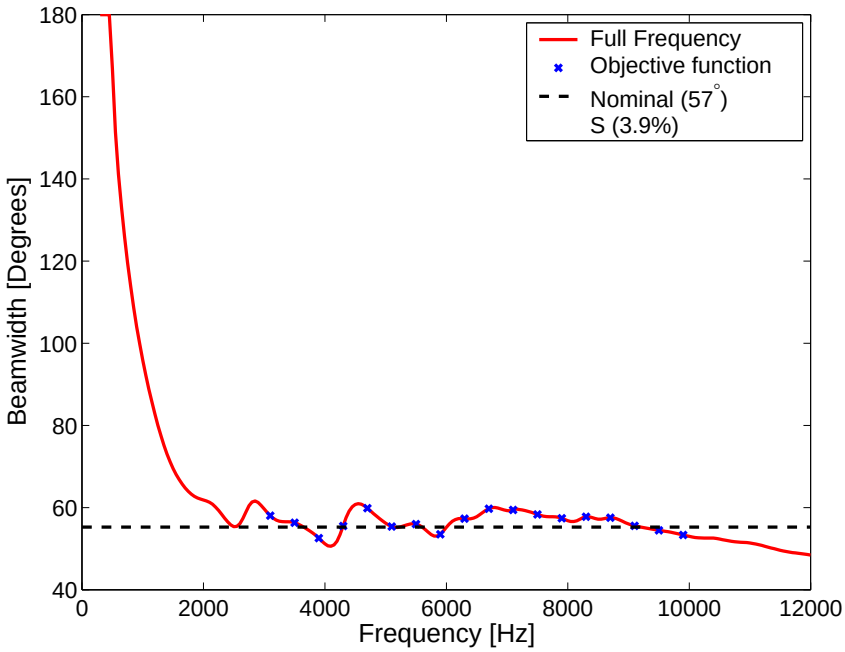
(θ=33°,F=50mm)

(b) Horn profile

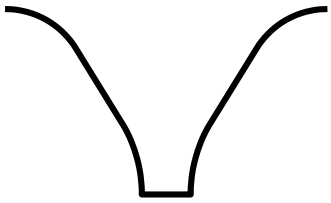
Figure 7.4: Results of the constrained optimisation of Φ_3 for the simple horn geometry with a 2 inch throat.

Referring to Figure 7.3, we can see that while the solution finds a minimum, if the constraint were removed the solution would most probably change and a smoother solution would be found. Relaxing the constraint, and using the DIRECT optimisation technique to find an unconstrained minimum of Φ_3, leads to $x(1) = 0.59$ ($\theta = 32°$) and $x(2) = 0.68$ ($F = 48$ mm). The mean value of the beamwidth (Φ_2) is 57° and the parameter S is 3.9%, only marginally smoother than the results for the constrained optimisation (4%). The beamwidth produced is shown in Figure 7.5 (a), and the horn profile in Figure 7.5

(b). This profile generates a beamwidth than is only marginally smoother than the previous constrained minimisation, and there is still a 'droop' in beamwidth at the higher frequencies.



(a) Beamwidth



(x(1)=0.59,x(2)=0.68)

(θ=32°,F=48mm)

(b) Horn profile

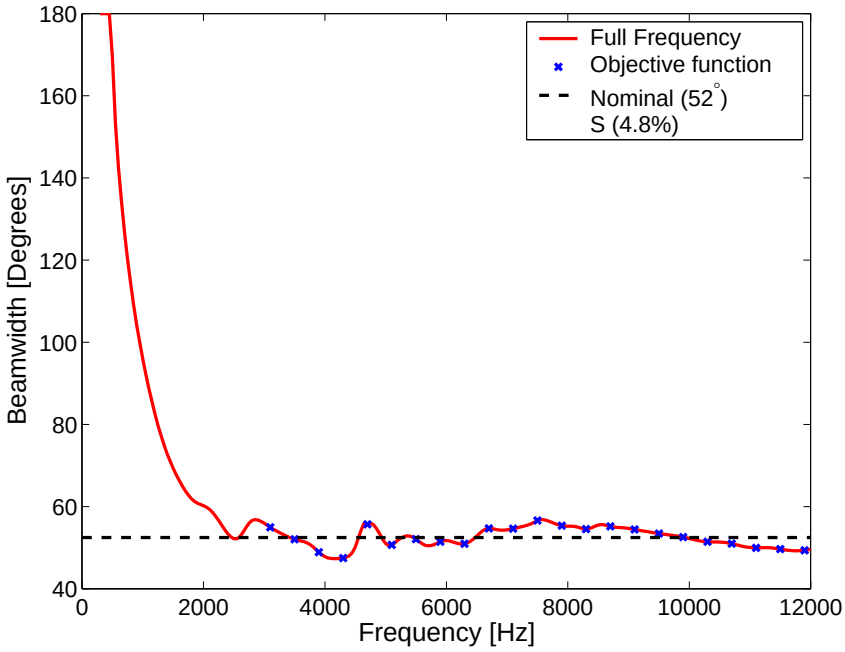
Figure 7.5: Results of the unconstrained optimisation of Φ_3 for the simple horn geometry with a 2 inch throat.

Experimental evidence from Figure 5.10 suggests that deficiencies in the compression driver may be blamed for poor experimental correlation above 10000 Hz. Notwithstanding the limitations of the numerical model when compared to experiment, this poor per-

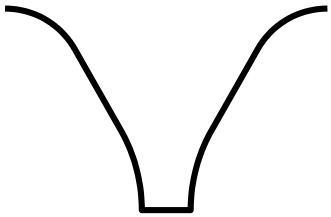
formance is inherent to the design of the horn and should be investigated further.

To try to find a solution that reduced the high frequency "droop", the frequency range over which the minimisation is performed is increased to 12000 Hz. The optimal solution for unconstrained minimisation with the new cost function is $x(1) = 0.48$ ($\theta = 30^\circ$) and $x(2) = 0.69$ ($F = 48$ mm), the mean value of the beamwidth $\Phi_2 = 53^\circ$ and the parameter S is 4.8%. The beamwidth produced is shown in Figure 7.6 (a), and the horn profile in Figure 7.6 (b). This profile generates a beamwidth that is less smooth than the previous constrained minimisation and does not achieve the desired nominal beamwidth. It does, however, minimise the high frequency "droop", and it is suggested that this cost function be implemented for future calculations wanting to achieve CBT performance.

In summary, this section shows that near constant beamwidth performance for horn loaded loudspeakers can be achieved using the optimisation techniques developed in this thesis, even with a very simple horn geometry parameterisation. It also finds that, in order to minimise a high frequency "droop", the frequencies over which the objective function is calculated should be extended to the highest possible frequency of interest.



(a) Beamwidth



(x(1)=0.59,x(2)=0.76)

(θ=30°,F=48mm)

(b) Horn profile

Figure 7.6: Results of the unconstrained optimisation of Φ_3 for the simple horn geometry with a 2 inch throat. The objective function function upper frequency limit is now 12000 *Hz, minimising high frequency "droop".*

7.3.2 One inch throat

The geometry described in the previous section has a fixed throat radius. In this section the same method of geometry parameterisation is investigated, but with a smaller 25 mm (1 inch) throat diameter, another standard dimension for compression drivers. Figure 7.7 shows that variation in geometry achieved by changing each parameter systematically between the lower limits, a mid range value and the upper limit.

The EGO optimisation technique is used to find the optimum geometry. Figure 7.8 shows the results of the optimisation. In this case, only 25 initial samples (cyan dots) are calculated, then 25 MAXVAR samples along the constraint (yellow dots) and 25 MOF samples (green dots). The minimum solution to the constrained optimisation is shown by the red dot. In Section 7.3.1 the DIRECT minimisation technique was used to find the unconstrained minimum function value. A simple modification to the objective functions to remove constraints in EGO allows for an efficient way to find the minimum of Φ_3, as shown by the black dot. Examination of the range of values achieved by this geometry, as shown in Figure 7.8, in comparison with Figure 7.3 shows that the 1 inch horn results are nowhere near as smooth as the 2 inch horn.

The optimal solution is $x(1) = 0.56$ ($\theta = 31^\circ$) and $x(2) = 0.46$ ($F = 35$ mm), the mean value of the beamwidth $\Phi_2 = 60.3^\circ$ and the parameter S is 12.5%. Figure 7.9 (a) shows the beamwidth of the constrained minimisation resulting from the geometry shown in Figure 7.9 (b). The desired value of nominal beamwidth is obtained at the expense of smoothness, and this design is not as smooth as the 2 inch horn results.

Relaxing the constraint on nominal beamwidth to investigate the smoothest solution possible shown by the black dot on Figure 7.8, $x(1) = 1.00$ ($\theta = 40^\circ$) and $x(2) = 0.43$ ($F = 33$ mm), the mean value of the beamwidth $\Phi_2 = 80^\circ$ and the parameter S is 6.9%. Figure 7.10 (a) shows the beamwidth of the unconstrained minimisation result, and the profile is shown in Figure 7.10 (b). The resulting nominal beamwidth is very large, with a large dip

below the nominal beamwidth at low frequencies. This performance in horns has been documented in the literature (Henricksen and Ureda, 1978). If the lowest frequency at which the objective function is calculated were raised above 4000 Hz, then this technique and geometry parameterisation could be used to design horns that perform well at very wide angles, at the expense of low frequency performance. They would not exhibit the characteristic CBT behaviour, and may have large variations in acoustic impedance due to reflections from the horn mouth.

The conclusions that can be drawn from this section are that the horn throat dimension with a simple flare into a conical horn is not an independent variable, and that a simple parameterisation is limited in its ability to generate optimal solutions. A parameterisation that includes the throat radius as a variable in the optimisation is required.

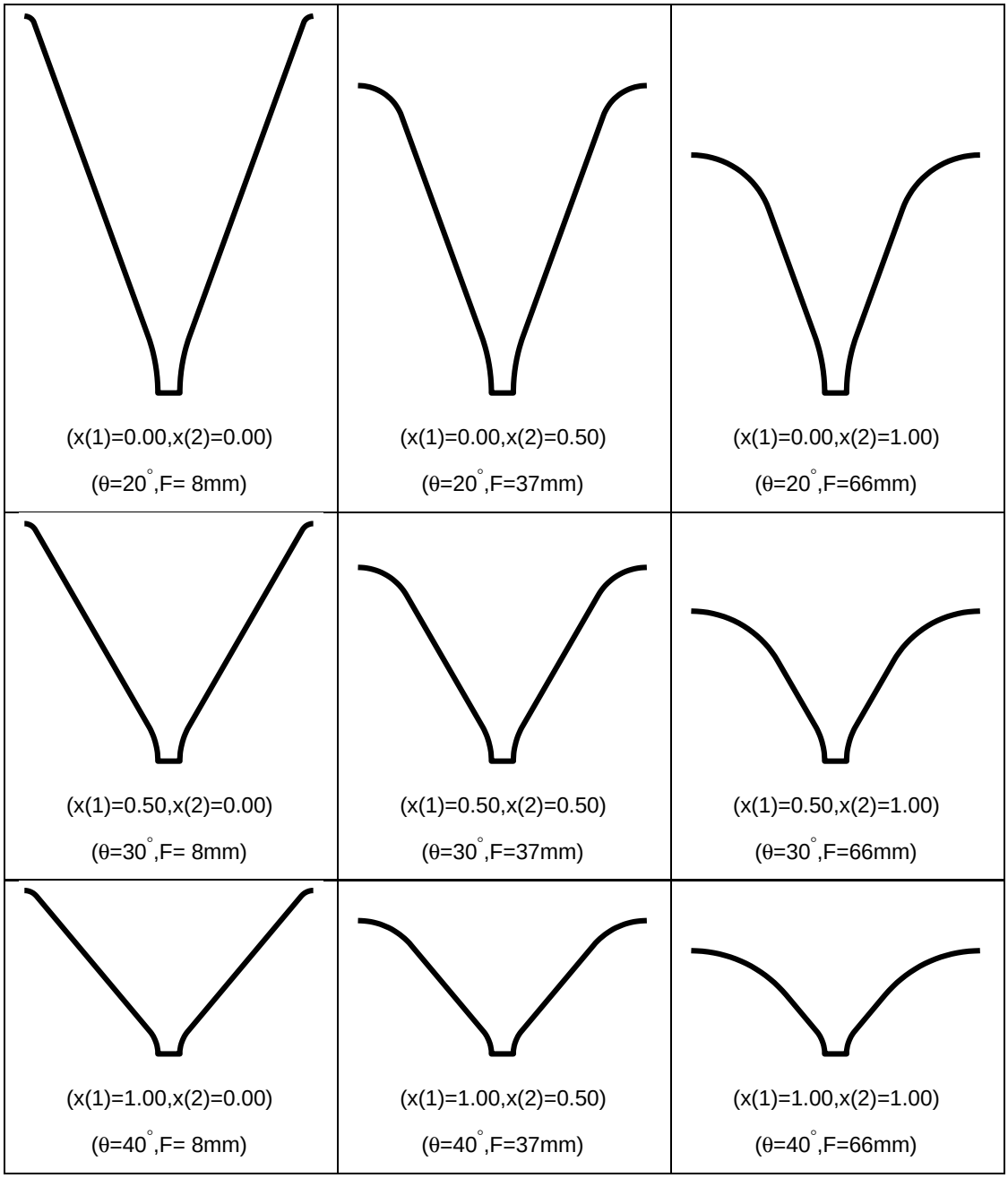


Figure 7.7: Variation in simple horn geometry with a 1 inch throat. Parameters vary between upper and lower bounds, $0 \leq x(1) \leq 1$ *and* $0 \leq x(2) \leq 1$.

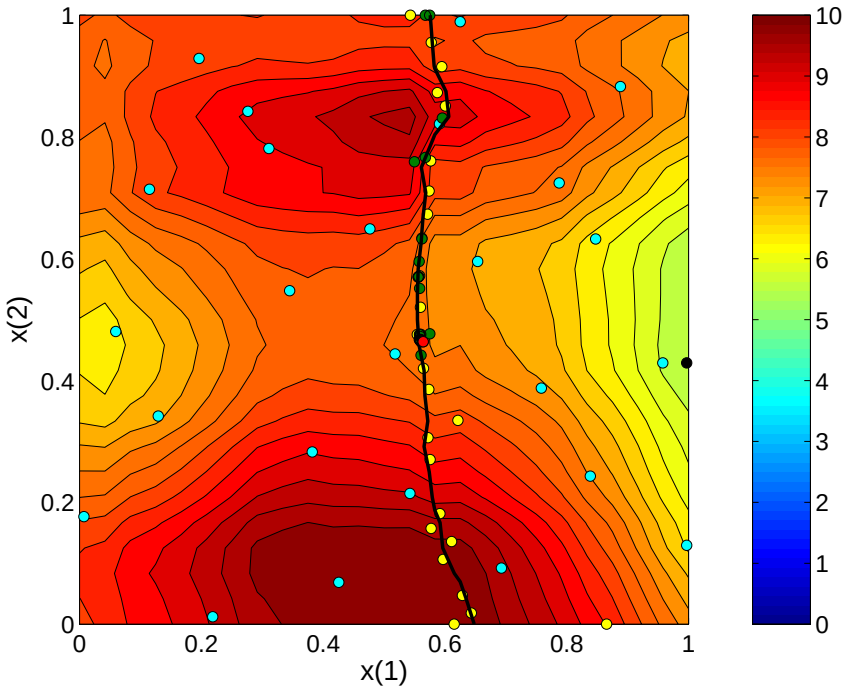
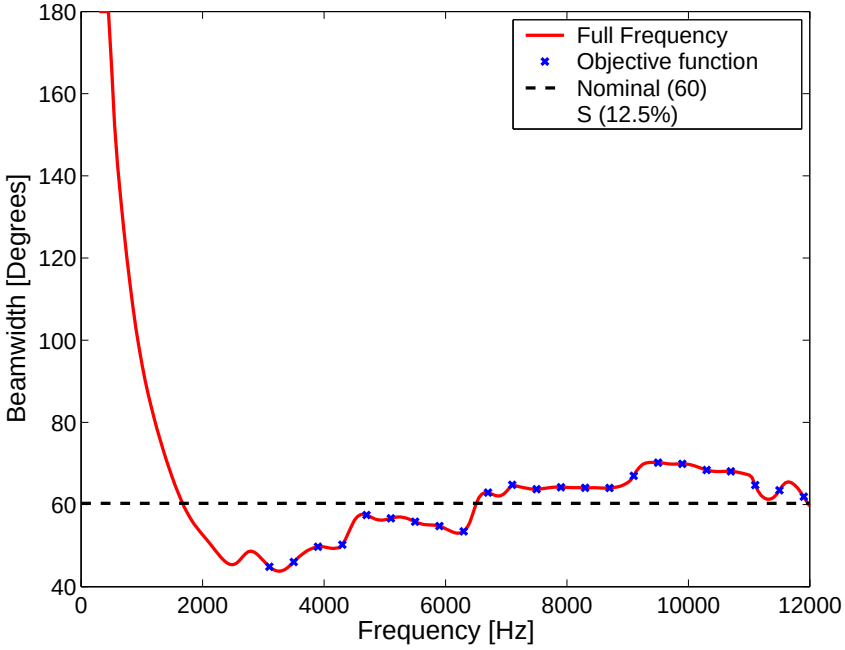
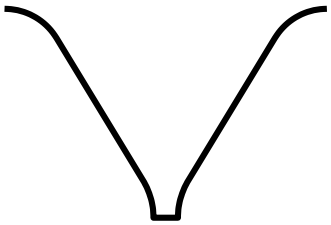


Figure 7.8: Optimisation trajectory for the simple horn geometry with a 1 inch throat. The cyan dots show the initial samples, yellow dots show the MAXVAR sampling, green dots the MOF sampling, the red dot shows the constrained global minimum with the black line showing the constraint. The black dot shows the unconstrained global minimum.



(a) Beamwidth

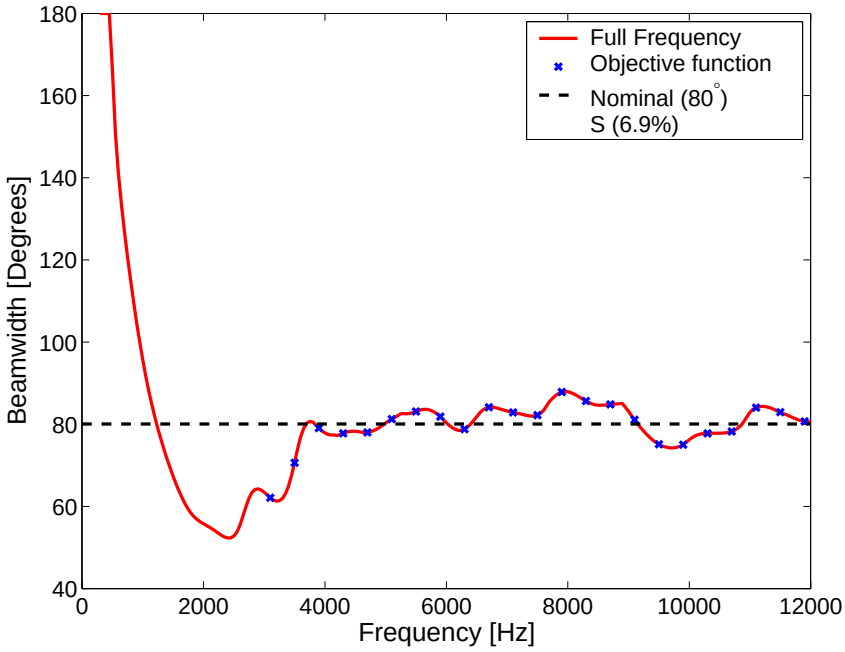


(x(1)=0.56,x(2)=0.46)

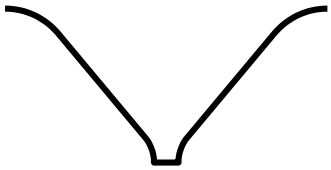
(θ=31°,F=35mm)

(b) Horn Profile

Figure 7.9: Results of the constrained optimisation of Φ_3 for the simple horn geometry with a 1 inch throat.



(a) Beamwidth



(x(1)=1.00,x(2)=0.43)

(θ=40°,F=33mm)

(b) Horn profile

Figure 7.10: Results of the unconstrained optimisation of Φ_3 for the simple horn geometry with a 1 inch throat.

7.3.3 Variable throat

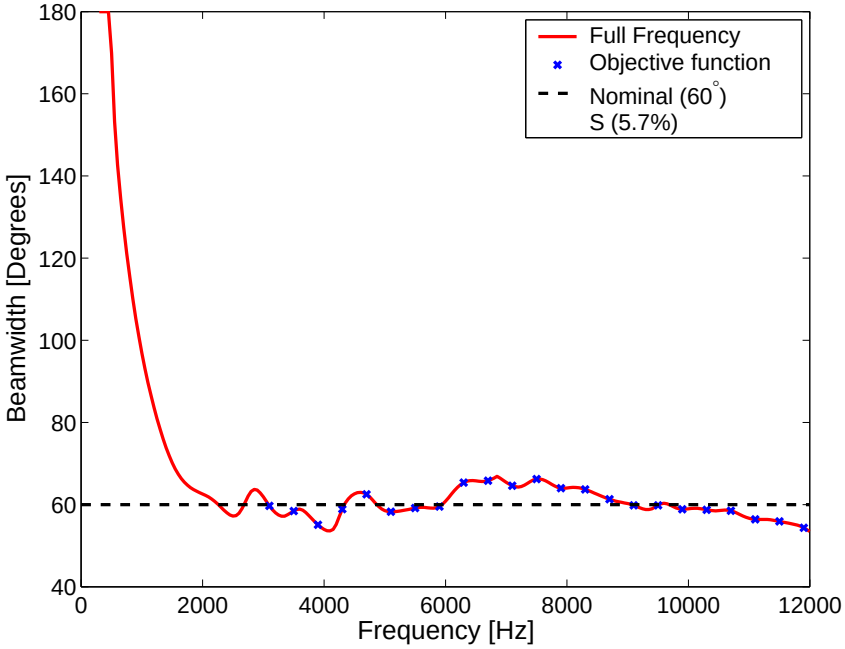
Optimisation of the simple flared horn with a variable throat radius is now investigated. Table 7.2 shows the upper and lower bounds of the variables used. The constrained EGO method finds an optimal solution $x(1) = 0.66$ ($\theta = 33^\circ$), $x(2) = 0.69$ ($F = 48$ mm) and $x(3) = 0.39$ ($D_1 = 22$ mm), the mean value of the beamwidth $\Phi_2 = 60^\circ$ and the parameter S is 5.7%. Figure 7.11 (a) shows the beamwidth of the constrained minimisation result resulting from the geometry shown in Figure 7.11 (b). The geometry found is similar to that of Figure 7.4 (b). This horn design appears to have a "drooping" beamwidth at high frequencies, and does not appear to achieve constant beamwidth behaviour. The horn geometry shown in Figure 7.4 (b) whilst not achieving constant beamwidth behaviour has achieved the aim of being as smooth as possible for a given nominal beamwidth.

Name	Variable	Lower	Upper
$x(1)$	θ	20°	40°
$x(2)$	$\frac{F}{R_m}$	5%	40%
$x(3)$	R_m	12.5 mm	37.5 mm

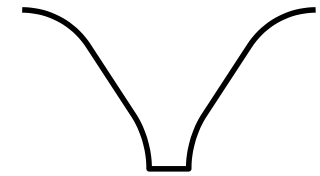
Table 7.2: Upper and lower bounds of parameters used to describe the simple flared horn geometry with a variable throat radius.

Performing an unconstrained minimisation on Φ_3 finds the optimal solution $x(1) = 0.14$ ($\theta = 23^\circ$), $x(2) = 0.52$ ($F = 38$ mm) and $x(3) = 0.94$ ($D_1 = 72$ mm), with the mean value of the beamwidth, $\Phi_2 = 37^\circ$ and the parameter $S = 3.4\%$. The beamwidth is shown in Figure 7.12 (a), and the profile in Figure 7.12 (b). This profile appears to closely approach constant beamwidth behaviour; however, the nominal beamwidth achieved is not that specified (60°). This suggests that either constant beamwidth behaviour may only occur over a limited range of nominal beamwidths, and that our goal of achieving constant beamwidth behaviour at a nominal angle of $60^\circ$is not possible, or that the geometry parameterisation used is too restrictive. Hence investigation into more flexible parameterisation is required for further progress toward constant beamwidth behaviour with a specified nominal beamwidth.

In summary, this section has used optimisation of a relatively simple 3 parameter horn geometry to find the horn shape that produces a relatively smooth beamwidth with a 60° nominal angle. However, this result, shown in Figure 7.11 shows a distinct "droop" at high frequencies, and cannot be considered to show constant beamwidth behaviour. It has also found a horn geometry that closely approached constant beamwidth behaviour at the expense of achieving a 37° nominal beamwidth instead of the specified 60° (Figure 7.12).



(a) Beamwidth

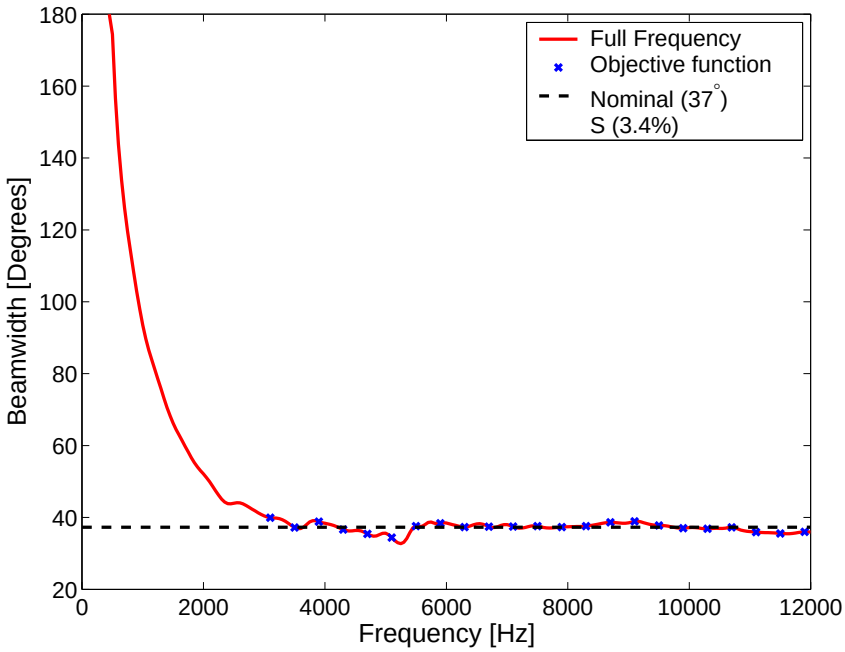


(x(1)=0.66,x(2)=0.69,x(3)=0.39)

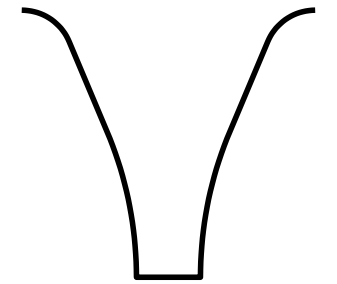
(θ=33°,Flange=48mm,D_1=45 mm)

(b) Horn profile

Figure 7.11: Results of the constrained optimisation of Φ_3 for the simple horn geometry with a variable throat dimension.



(a) Beamwidth



(x(1)=0.14,x(2)=0.52,x(3)=0.94)

(θ=23°,Flange=38mm,D_1=72 mm)

(b) Horn profile

Figure 7.12: Results of the unconstrained optimisation of Φ_3 for the simple horn geometry with a variable throat dimension.

7.4 Bézier horn

Bézier curves were introduced in Section 6.9 to provide more flexibility in specifying the velocity profile over the surface of a sphere. The Bézier spline is specified by two vectors, with the curve tangent to the head of each vector, and the "strength" of attachment to the vector determined by the length of the vector. They are described in detail in Appendix D.

The horn geometry has been parameterised using Bézier curves. The length ($L = 235$ mm), mouth radius (R_m = 82.5 mm) and throat radius ($R_t = 25$ mm) are held constant. The head of one Bézier vector is at $(R_t, 0)$, with the tail placed at fraction $x(1)$ between the head and (R_m, L). Similarly the head of the second Bézier vector is placed at (R_t, L), with the tail placed at fraction $x(2)$ between the head and (R_m, L). The parameters $x(1)$ and $x(2)$ control the shape of the horn. Figure 7.13 shows the variation in geometry achieved by changing the parameters systematically.

The EGO constrained minimisation technique was unable to find a solution with a nominal beamwidth $\mathcal{B}_{nom} = 60^\circ$. This is because the current geometry parameterisation is physically unable to produce a beamwidth of this magnitude. The EGO unconstrained minimisation of Φ_3 was able to find a smooth solution with a nominal beamwidth of 53° and the parameter S is 5.7%, shown in Figure 7.14 (a), from the profile $x(1) = 0.28$ and $x(2) = 0.29$, shown in Figure 7.14 (b). This geometry profile does not produce a particularly smooth beamwidth compared to the unconstrained minimisation of a simple flared horn. This parameterisation does not appear to exhibit constant beamwidth behaviour, with a large undershoot of the nominal beamwidth at low frequencies. Other parameterisations with different control over the geometry profiles are investigated in the next section.

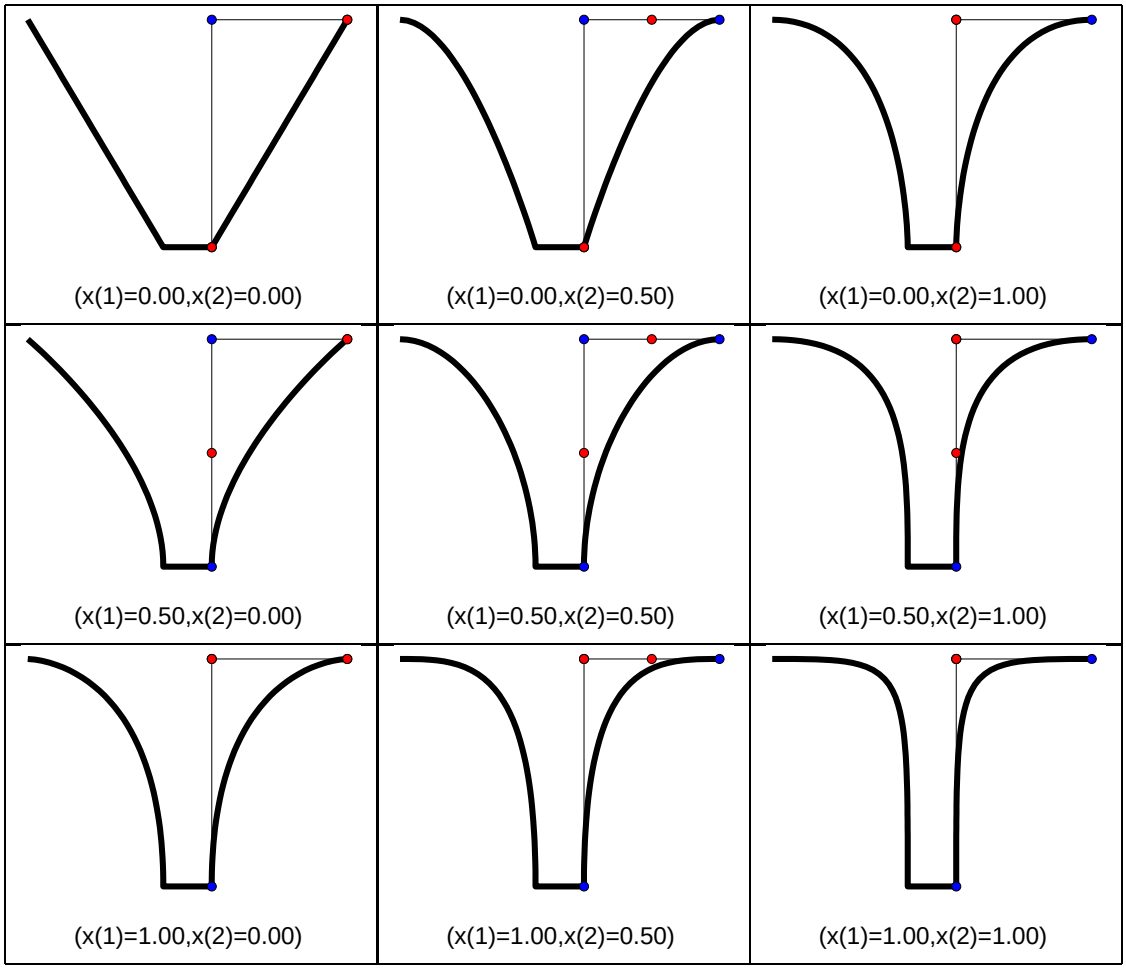
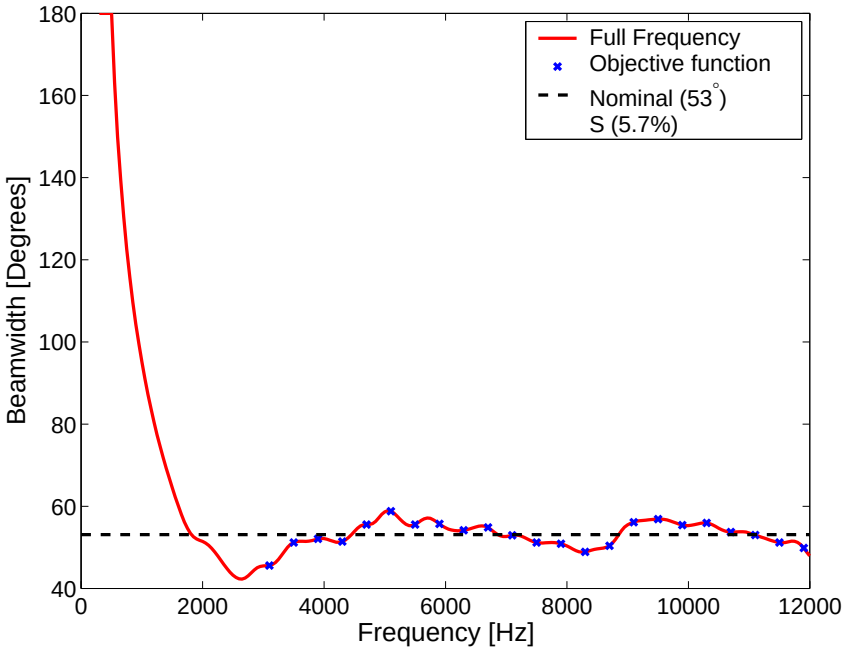
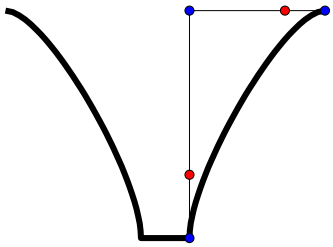


Figure 7.13: Variation in Bézier horn geometry with a 2 inch throat and horn length 235 *mm. Parameters vary between upper and lower bounds,* $0 \leq x(1) \leq 1$ *and* $0 \leq x(2) \leq 1$.



(a) Beamwidth



(x(1)=0.28,x(2)=0.29)

(b) Horn profile

Figure 7.14: Results of the unconstrained optimisation of Φ_3 for the Bézier horn geometry with a 2 inch throat and horn length 235 *mm.*

7.5 Spline based horns

As a simple Bézier curve is unable to produce fine enough control over the shape of the geometry profile, the use of Bézier splines to describe the horn geometry is investigated. Bézier splines, as implemented in the ANSYS (Kohnke, 2001) APDL command `bspline`, allow any number of points to control the curve, as well as tangency conditions at each end point. Much greater variation in geometry can be achieved than was possible using the simple parameterisations described in previous sections.

7.5.1 Simple spline horn

This section investigates a relatively simple spline parameterisation. The horn throat radius was held constant at $D_t = 2R_t = 50$ mm (2 inches), the length of the horn fixed at L =235 mm and the radius of the mouth fixed at $D_m = 2R_m$ = 165 mm.

The start and end points of a parametric cubic spline are given by the points $(R_t, 0)$ and (R_m, L) respectively. Tangency is enforced at the start and end points in the directions $(0,1)$ and $(-1,0)$ respectively. A point in the spline curve is allowed to move in a rectangular box aligned along the start and end points. One parameter, $x(1)$, controls the major axis position, the other parameter $x(2)$ controls the minor axis. Figure 7.15 shows the range of geometries possible with this parameterisation, along with the position and shape of the box.

Figure 7.16 shows the results of the optimisation, with 25 initial samples (cyan dots), 25 MAXVAR samples along the constraint (yellow dots) and 25 MOF samples (green dots). The solution to the constrained optimisation of Φ_3 is shown by the red dot, $x(1) = 0.01$ and $x(2) = 0.81$, and the solution of the unconstrained optimisation of Φ_3 shown by the black dot, $x(1) = 1$ and $x(2) = 1$.

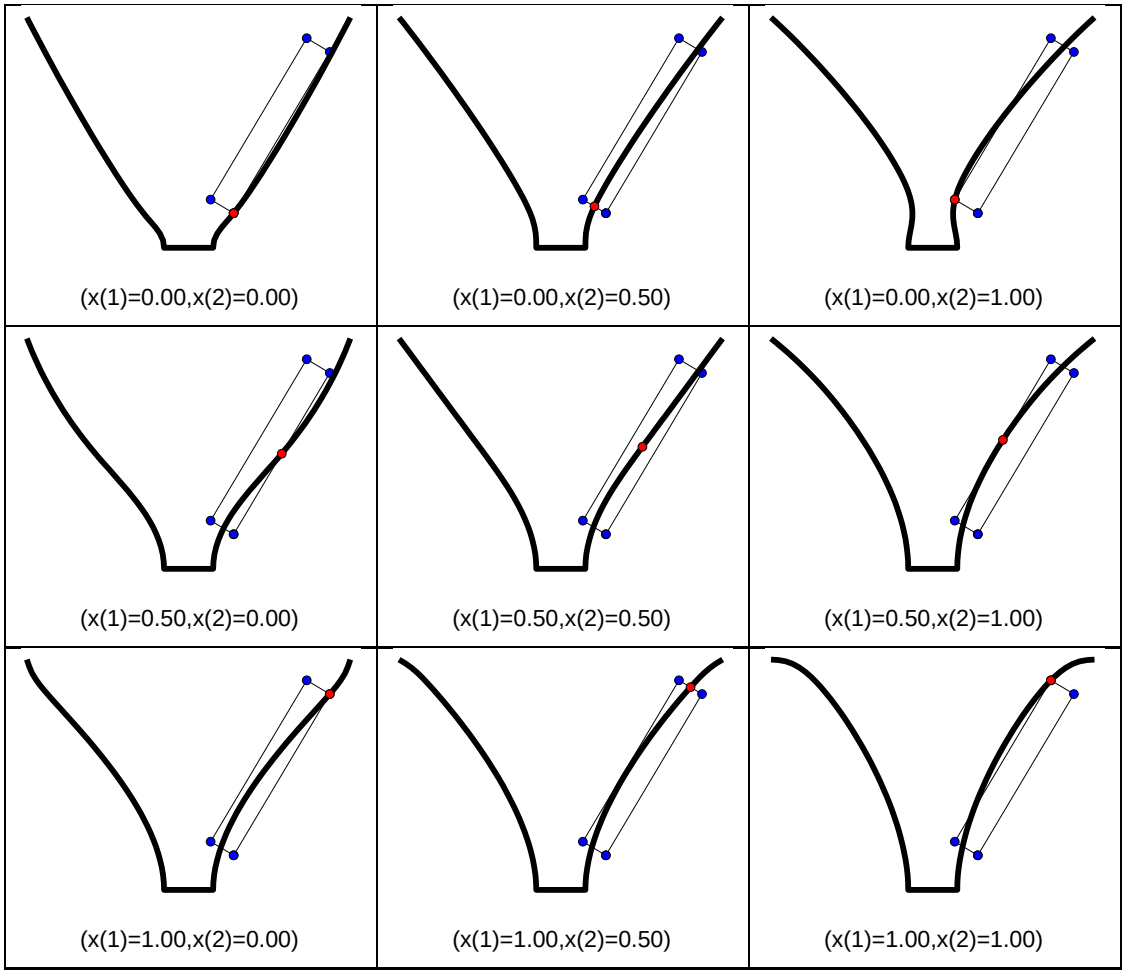


Figure 7.15: Variation in simple spline horn geometry with a 2 inch throat and horn length 235 *mm. Parameters vary between upper and lower bounds,* $0 \leq x(1) \leq 1$ *and* $0 \leq x(2) \leq 1$.

Figure 7.17 shows the beamwidth and geometry profile of the constrained minimisation of Φ_3, and Figure 7.9 shows the same for the unconstrained minimisation. Neither beamwidth is particularly smooth, and this geometry parameterisation is not particularly useful.

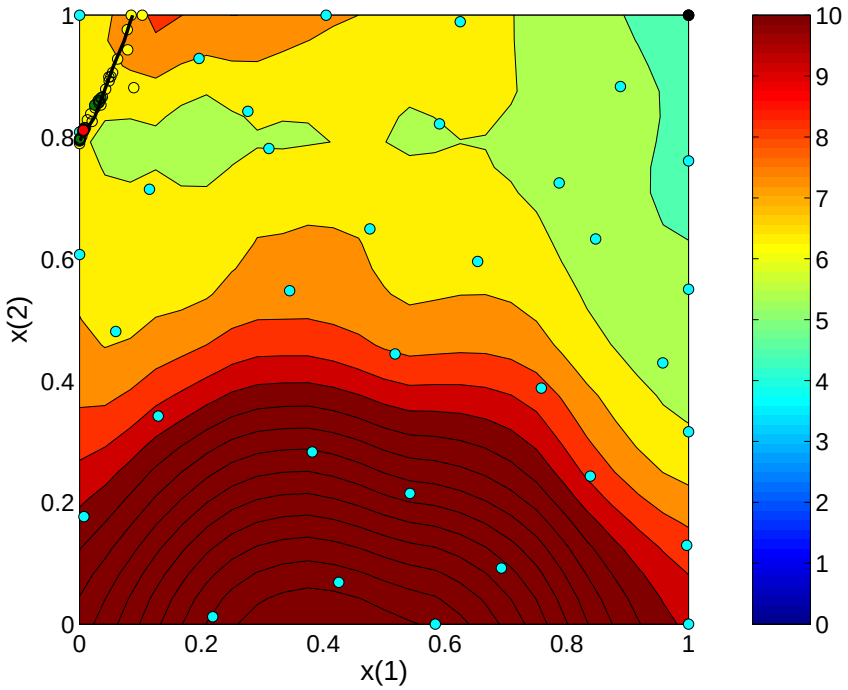
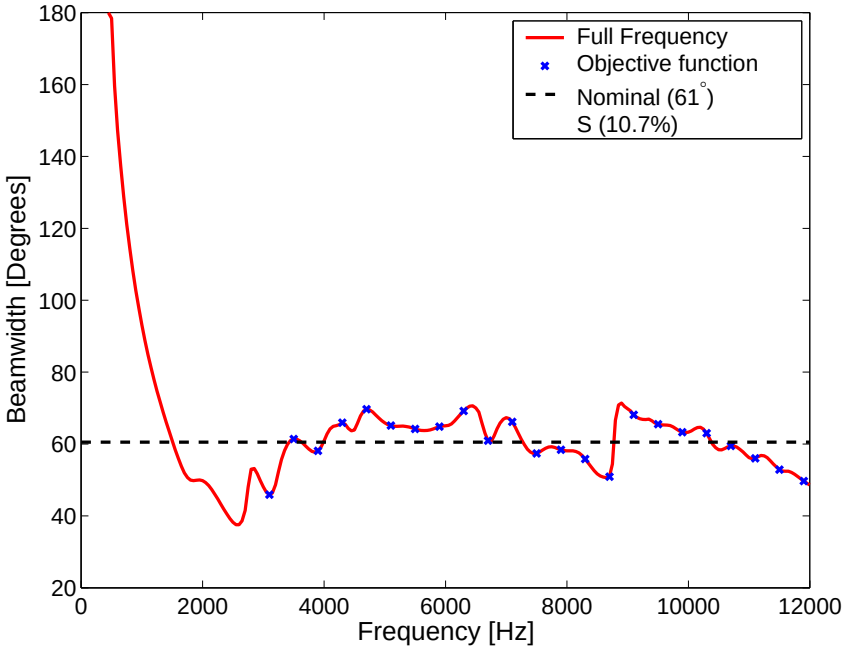
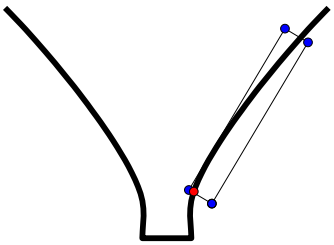


Figure 7.16: Optimisation trajectory for the simple spline horn geometry with a 2 inch throat and horn length 235 *mm. The cyan dots show the initial samples, yellow dots show the MAXVAR sampling, green dots the MOF sampling and the red dot shows the constrained global minimum with the black line showing the constraint. The black dot shows the constrained global minimum.*



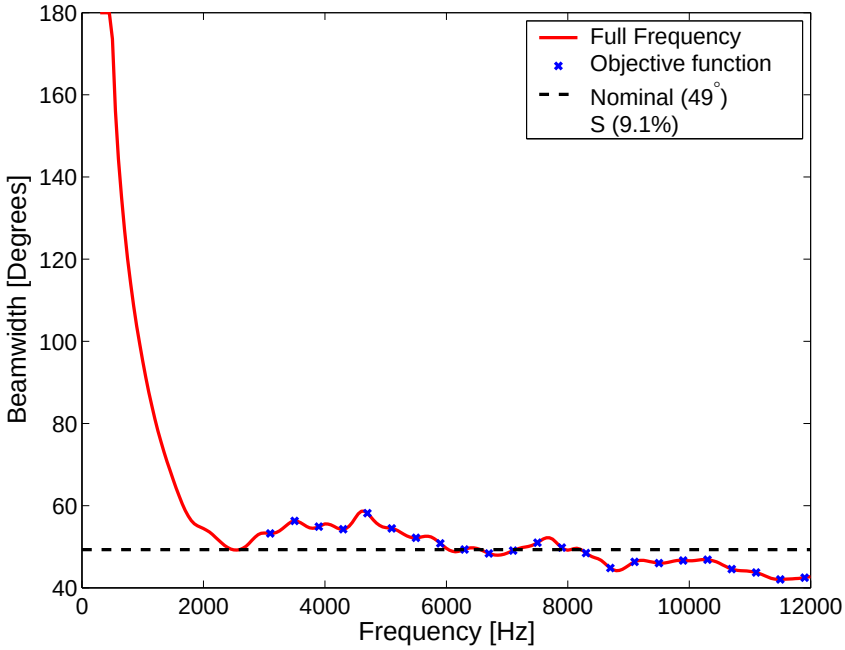
(a) Beamwidth



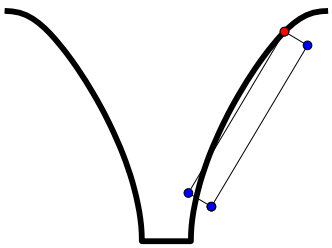
(x(1)=0.01,x(2)=0.81)

(b) Horn profile

Figure 7.17: Results of the constrained optimisation of Φ_3 for the simple spline horn geometry with a 2 inch throat and horn length 235 *mm.*



(a) Beamwidth



(x(1)=1.00,x(2)=1.00)

(b) Horn profile

Figure 7.18: Results of the unconstrained optimisation of Φ_3 for the simple spline horn geometry with a 2 inch throat and horn length 235 *mm.*

7.5.2 Complex spline horn

A more complex spline based horn geometry is considered in this section. The horn throat radius was held constant at $D_t = 2R_t = 50$ mm (2 inches), the length of the horn fixed at L =235 mm and the radius of the mouth fixed at $D_m = 2R_m$ = 165 mm. A spline is fit using the same start, end and tangency conditions as those in Section 7.5 and two control points also control the shape of the curve. The position of each point is controlled by lines starting $\frac{1}{8}$ and $\frac{2}{3}$ of the distance between the start and end point, and extending 80 mm vertically from each point. The fraction $x(1)$ controls the position of the first control point between its start and end points, and $x(2)$ the position of the second control point. Figure 7.19 shows the range of geometries possible with this parameterisation, with the control points shown as red dots and the start and end points by blue dots.

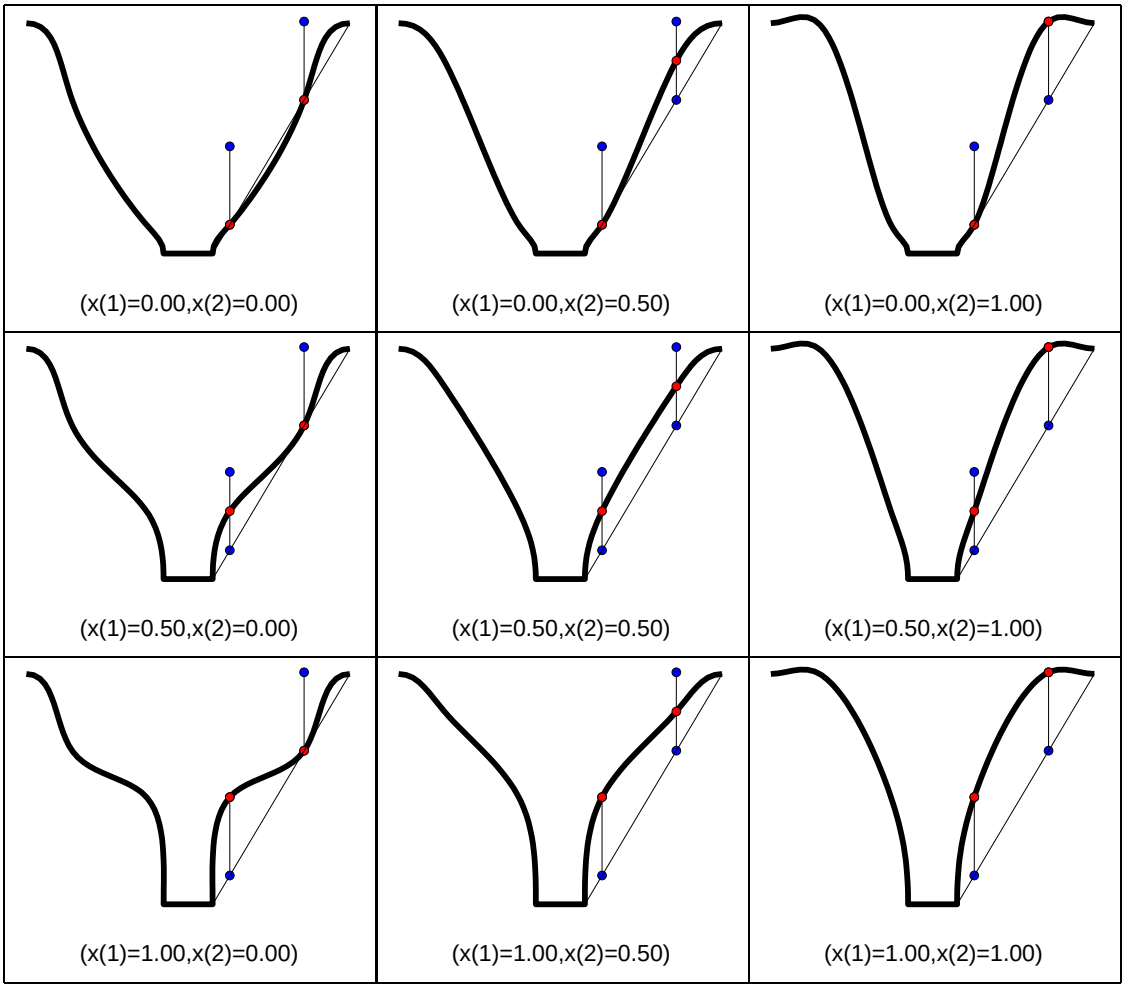
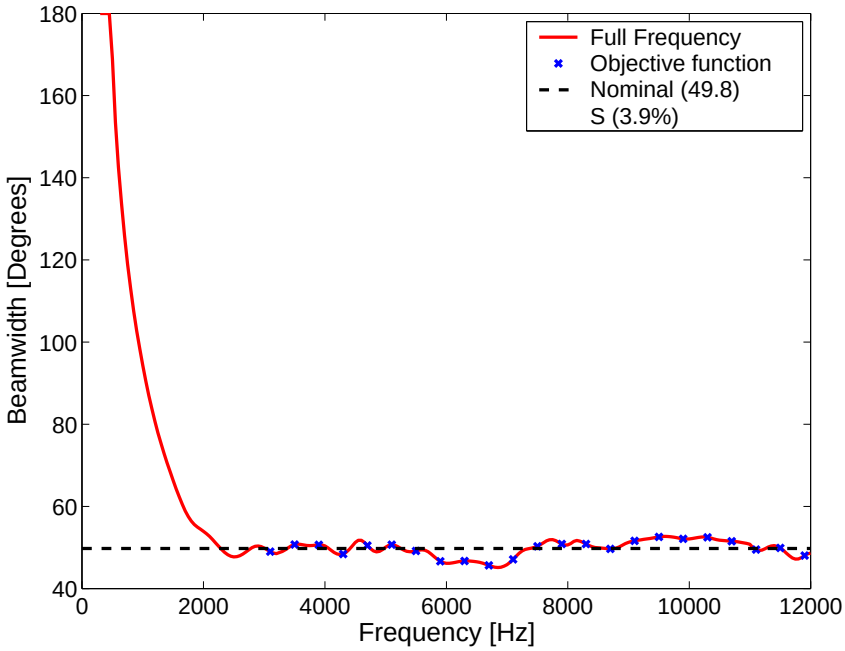


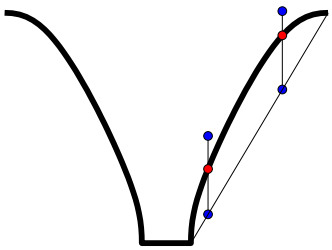
Figure 7.19: Variation in complex spline horn geometry for with a 2 inch throat and horn length 235 *mm. Parameters vary between upper and lower bounds,* $0 \leq x(1) \leq 1$ *and* $0 \leq x(2) \leq 1$.

EGO optimisation of this geometry is not able to find a constrained solution with nominal beamwidth $\mathcal{B}_{nom} = 60^\circ$. Figure 7.20 (a) shows the beamwidth found by minimising Φ_3 without imposing beamwidth constraints. This horn has an acceptable constant beamwidth behaviour, although it does not reach the desired nominal beamwidth. To change the nominal beamwidth it appears as though both the length of the horn and the throat radius are required as parameters. The horn profile that generates this beamwidth is given in Figure 7.20 (b), with $x(1) = 0.58$ and $x(2) = 0.69$, the nominal beamwidth 49.8° with parameter $S = 3.9\%$. The profile appears similar to the simple flared horn designs (c.f. Figure 7.6 (b)), with the spline points controlling the entry and exit flare rates.

To examine the hypothesis that the horn length has an effect on the nominal beamwidth, a number of unconstrained optimisations of Φ_3 were performed using the complex spline horn geometry using different values of horn length, L. Figures 7.21,7.22 and 7.23 show the results for $L = 260$ mm, $L = 285$ mm and $L = 210$ mm respectively. Varying the length does not seem to have a consistent effect on the nominal beamwidth, with the 260 mm and 280 mm horns having almost identical nominal beamwidths, although they both have different values of the smoothness parameter S. The $L = 210$ mm horn profile, shown in Figure 7.23 (b), has an optimum with both control points at close to the limits. This forces the flange to curve over, or become re-entrant. This phenomena is examined further in Section 7.5.4.



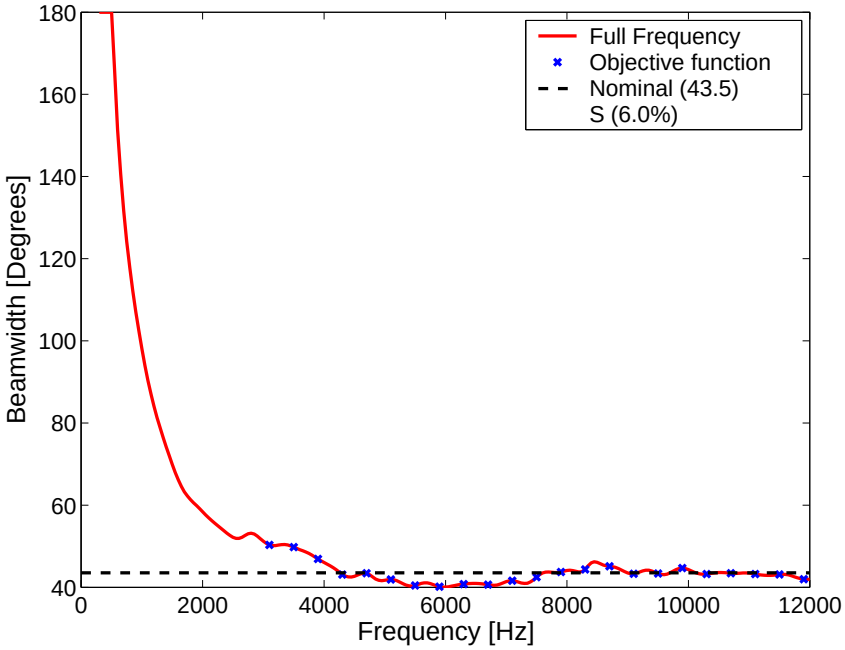
(a) Beamwidth



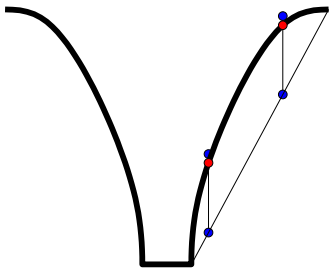
(x(1)=0.58,x(2)=0.69)

(b) Horn profile

Figure 7.20: Results of the unconstrained optimisation of Φ_3 for the complex spline horn geometry with a 2 inch throat and horn length 235 *mm.*



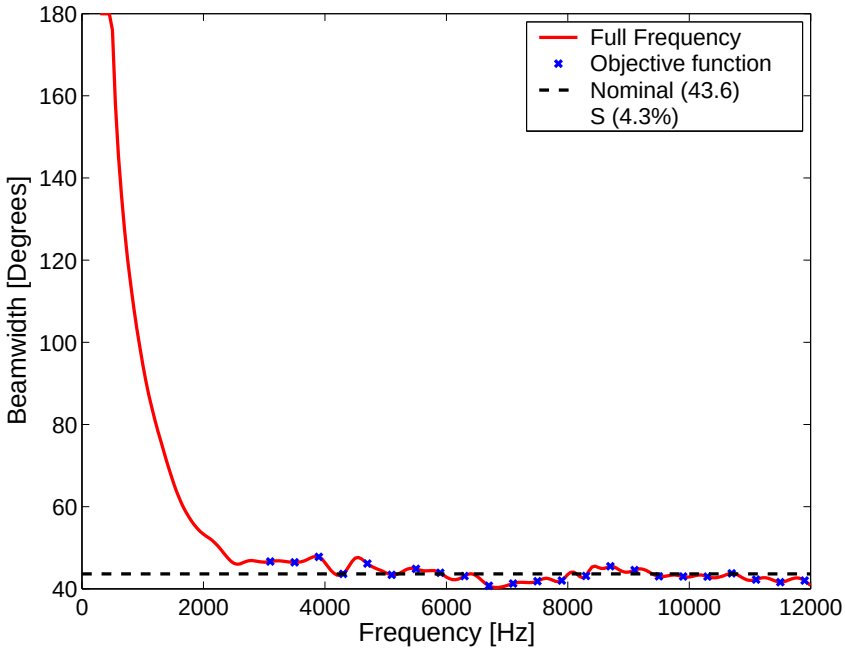
(a) Beamwidth



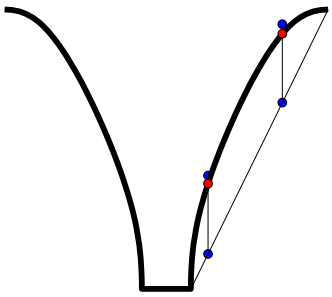
(x(1)=0.89,x(2)=0.88)

(b) Horn profile

Figure 7.21: Results of the unconstrained optimisation of Φ_3 for the complex spline horn geometry with a 2 inch throat and horn length 260 *mm.*



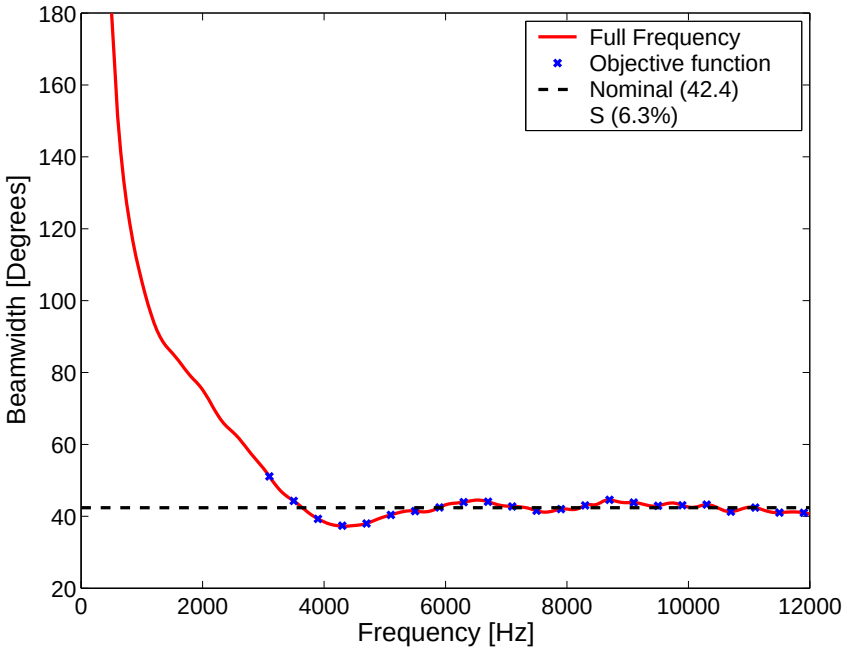
(a) Beamwidth



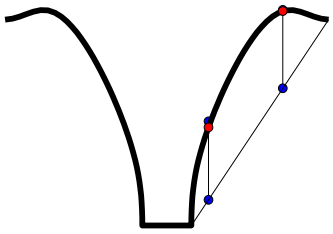
(x(1)=0.90,x(2)=0.88)

(b) Horn profile

Figure 7.22: Results of the unconstrained optimisation of Φ_3 *for the complex spline horn geometry with a 2 inch throat and horn length* 285 *mm.*



(a) Beamwidth



(x(1)=0.92,x(2)=0.98)

(b) Horn profile

Figure 7.23: Results of the unconstrained optimisation of Φ_3 for the complex spline horn geometry with a 2 inch throat and horn length 210 *mm.*

7.5.3 4 parameter optimisation

The complex spline based parameterisation appears to be a promising approach to finding an optimum horn shape. The existing parameterisation is not flexible enough to find a optimum that is constrained to the nominal beamwidth. To make the shape more flexible, both the horn length (L) and throat radius (R_t) are made parameters, along with the existing spline control points. The length of the control lines is increased from 80 mm to 100 mm to allow greater movement. Table 7.3 gives the upper and lower bounds on the variables used in this simulation.

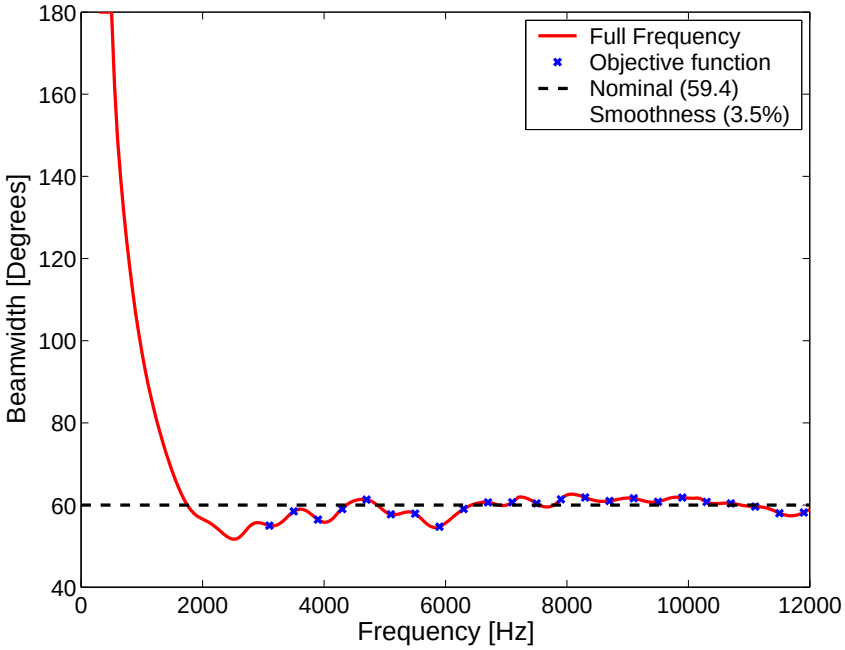
Name	Variable	Lower	Upper
$x(1)$	Spline control point 1	0	1
$x(2)$	Spline control point 2	0	1
$x(3)$	R_t	12.7 mm	82.5 mm
$x(4)$	L	200 mm	400 mm

Table 7.3: Upper and lower bounds of parameters used to describe the 4 parameter horn geometry.

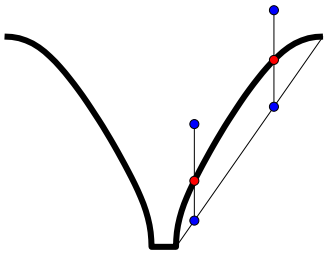
The application of the constrained EGO optimisation technique is able to find an optimum solution, although with the MAXVAR / MOF sampling ISC there is no way of deciding if a global optimum has been found. It is possible that some form of sampling regime (the auxiliary optimisation performed by the EGO optimisation method, see Section 7.2) that switches automatically between MAXVAR and MOF could be implemented, although it may be preferable to work on improving the constrained optimisation of the EI ISC (Section 8.3).

Figure 7.24 (a) shows the optimum beamwidth (59.4°) with a smoothness of 3.5%. This is generated by the profile shown in Figure 7.24 (b) with $x(1) = 0.40$, $x(2) = 0.47$, $R_t = 12.7$ mm and $L = 218$ mm. This design has the smallest possible R_t and L, making the included angle of the horn as wide as possible. The beamwidth dips beneath the nominal beamwidth at low frequencies, indicating that this design, while reasonably smooth, only exhibits marginally constant beamwidth behaviour. A suggested area of future research is

the development of cost functions that do not choose a low frequency cut off, instead rewarding constant beamwidth behaviour that smoothly approaches the nominal beamwidth and penalising other behaviour (Section 8.3).



(a) Beamwidth



(x(1)=0.41,x(2)=0.48,x(3)=1.00,x(4)=0.09)

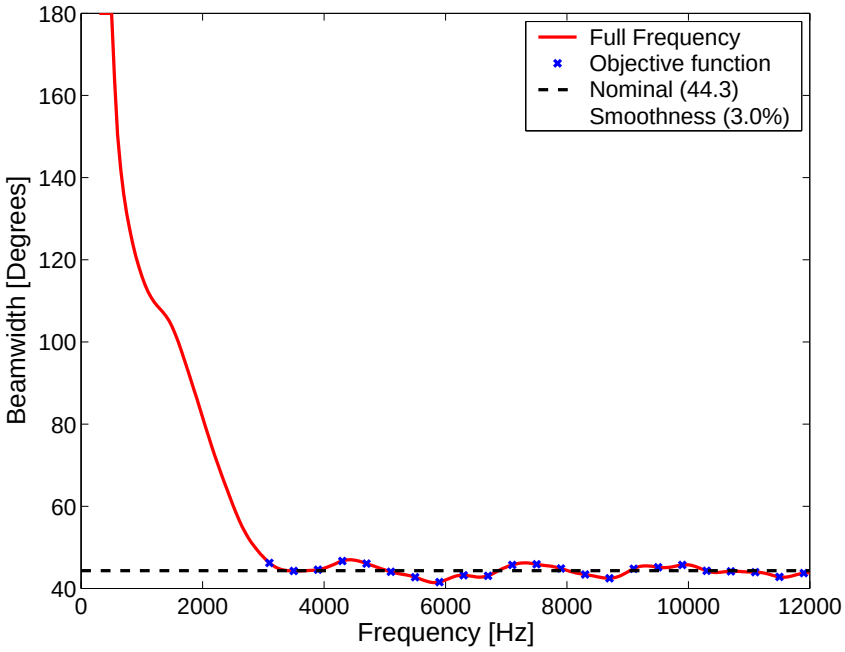
(R_t=12.7mm,L=218mm)

(b) Horn profile

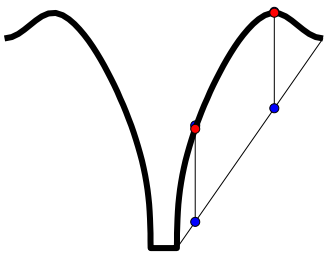
Figure 7.24: Results of the constrained optimisation of Φ_3 for the complex spline horn geometry with a variable throat dimension and horn length.

Optimising Φ_3 and removing the constant beamwidth constraint produces a re-entrant flange condition, although this solution may not be the true global minimum because the Infill Sampling Criteria (ISC) used is the Minimum Objective Function (MOF), and

this does not guarantee that a global minimum will be found. Switching to the Expected Improvement (EI) ISC (Equation B.4) will make the solution more robust, and is recommended for future unconstrained optimisations. The re-entrant flange is obtained when the control line is pushed out beyond the mouth of the horn, and the geometry is able to form a smooth curve. It appears to effectively make the mouth of the horn smaller. This condition requires further investigation to decide whether it is a true optimum, or attributed to poor geometry parameterisation.



(a) Beamwidth



(x(1)=0.97,x(2)=0.99,x(3)=0.92,x(4)=0.09)

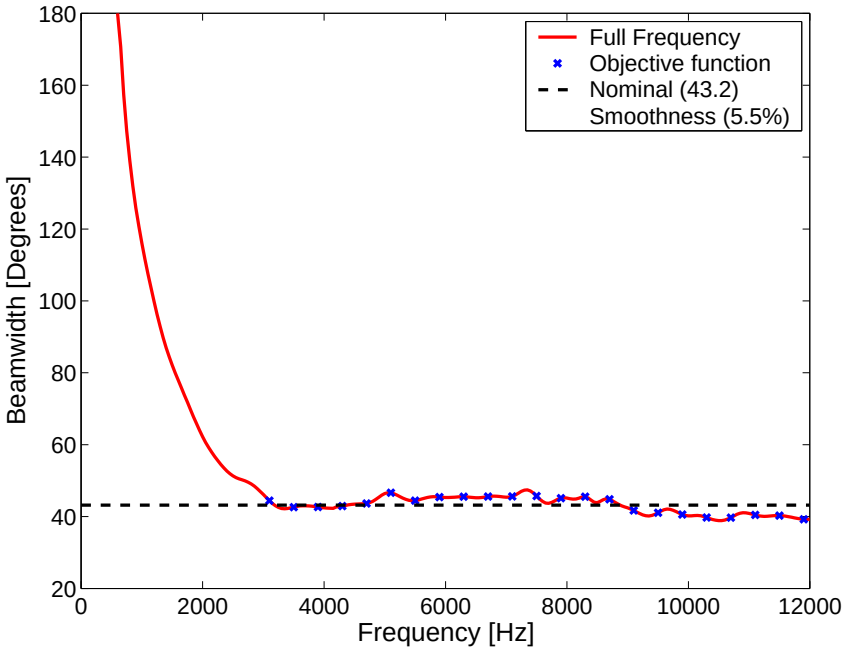
(R_t=13.6mm,L=217mm)

(b) Horn profile

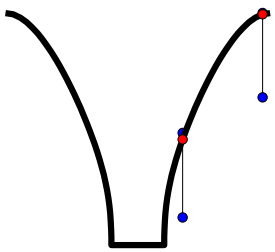
Figure 7.25: Results of the unconstrained optimisation of Φ_3 for the complex spline horn geometry with a variable throat dimension and horn length.

7.5.4 Re-entrant flange

The re-entrant flange is an interesting phenomena which is seen in the previous section in Figure 7.25 (b) as well as in the $L = 210$ mm horn in Figure 7.23 (b). The flange is curved over (re-entrant) outside the horn mouth, effectively changing the horn mouth dimension. To see if the flange is important to the performance of the horn, the re-entrant part of the flange was removed from the model used to generate Figure 7.23 (b), leaving the mouth dimension $D_m = 2R_m = 251$ mm. The horn geometry is shown in Figure 7.26 (b) and resulting the beamwidth shown in Figure 7.26 (a). The superior performance of this configuration ($S = 5.5\%$ over $S = 6.3\%$) , without any extra optimisation, implies that the re-entrant flange should be suppressed from future geometric parameterisations.



(a) Beamwidth



(x(1)=0.92,x(2)=0.98)

(b) Horn profile

Figure 7.26: Results of the unconstrained optimisation of Φ_3 for the complex spline horn geometry with a 2 inch throat and horn length 210 mm. The re-entrant part of the flange has been removed, which results in superior performance ($S = 5.5\%$) to that seem with the flange in Figure 7.23 ($S = 6.3\%$).

7.5.5 Smooth Expected Improvement optimisation

An issue with the previous unconstrained optimisations is that the standard deviation of the beamwidth (Φ_3) has been optimised, instead of a measure of variance normalised by the mean value of the beamwidth (Φ_2), such as the parameter S (Equation 7.7). For the original, constrained optimisation, finding a minimum of Φ_3 is equivalent to minimising the parameter S, as Φ_2 , which S depends on, is held fixed by the constraint. However for an unconstrained optimisation, the minimum of Φ_3 is not necessarily the minimum of S, as Φ_2 can now vary. The differences in the minima found by optimising S instead of Φ_3 are not expected to be great, but parameter S gives a better measure of constant beamwidth horn behaviour, and should be adopted for future calculations.

Another issue involves the optimisation technique used. For constrained EGO optimisation it was found in Section 6.8 that it was necessary to use a number of iterations using the constrained MAXVAR ISC to efficiently search the constrained objective function space, followed by more iterations using the constrained MOF ISC to find the global minimum. Using this method to find a global minimum for an unconstrained problem is inefficient as the search for a global optimum performed with two separate searches. It was found in Section 6.8 that the Expected Improvement (EI) ISC is able to efficiently balance local and global searches for unconstrained minimisation, and this approach is adopted here.

Changing the parameterisation of the horn, so that the end point of the lines that govern the position of the spline control points are now made a function of the horn length rather than being fixed, eliminates the potential of a re-entrant flange. The fractions $x(1)$ and $x(2)$ control the position of the control points between the start and end points of the control lines. The positions of the lines are still $\frac{1}{8}$ and $\frac{2}{3}$ of the distance between the start and end points, but the end point of the first control line is $\frac{1}{2}$ of the axial distance between the start point and the horn mouth, which can be seen in Figure 7.28 (b). The end point of the second control line is in line with the horn mouth. This scales the control lines with the length of the horn.

The results of the EGO EI optimisation of S for a horn with fixed length $L = 260$ mm and 2 inch throat appear in Figure 7.27. The 25 cyan dots show where the initial points are sampled, and the 25 green dots show the sample points chosen by the EI ISC, balancing both local and global optimisation. A convergence to the global minimum can be seen with repeated sampling (many green dots) around the global minimum (red dot) at $x(1) = 0.49$ and $x(2) = 0.69$. We can now be reasonably certain that the global minimum of an unconstrained 2 parameter optimisation of S can be found within 50 function evaluations. The beamwidth is shown in Figure 7.28 (a) with a nominal beamwidth of 47.1° with parameter $S = 3.1\%$. The profile is shown in Figure 7.28 (b). This calculations should compare directly with that in Figure 7.21. Both parameterisations are capable of producing the same geometry, but the EGO optimisation with EI ISC finds a better solution, and should be adopted for future unconstrained optimisations.

Extending the geometry to include a third control point between the two existing points gives finer control over the shape of the horn. The fractions $x(1)$, $x(2)$ and $x(3)$ control the positions of the three control points between the start and end points of the control lines. The unconstrained EGO with EI ISC is able to find a global minimum, $x(1) = 0.48$, $x(2) = 0.43$ and $x(3) = 0.70$, shown as the red dot in Figure 7.29. The 35 cyan dots show the initial samples, and the 65 green dots show the samples chosen by the EI ISC. There is a large amount of sampling around the red dot, showing dense sampling and likely convergence to a global minimum.

The beamwidth produced by this optimal solution is shown in Figure 7.30 (a). The nominal beamwidth is 46.8° with a parameter $S = 2.8\%$. The profile, shown in Figure 7.30b is remarkably similar to the 2 parameter version shown in Figure 7.28 (b). The smoothness of the beamwidth indicates that constant beamwidth behaviour likely to be achieved with this design.

Figure 7.31 (a) and (b) show the absolute and percentage differences between the radial profiles of the 2 and 3 parameter complex spline horns, as shown in Figure 7.28 (b) and

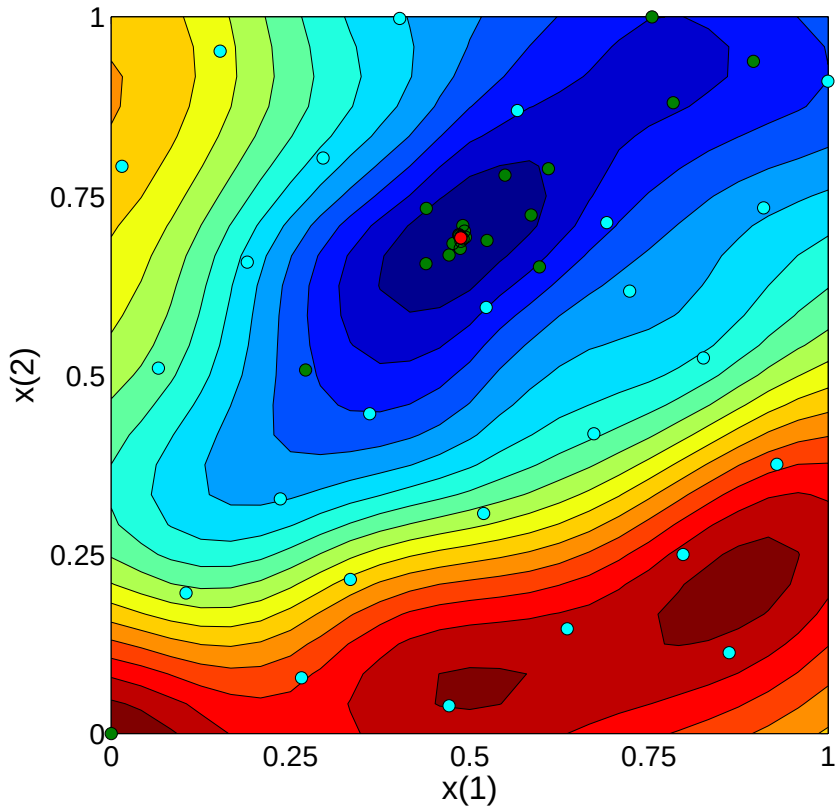


Figure 7.27: Optimisation trajectory for the 2 parameter complex spline horn geometry with a 2 inch throat and $L = 260$ mm. The cyan dots show the initial samples, green dots the EI sampling and the red dot shows the unconstrained global minimum.

.

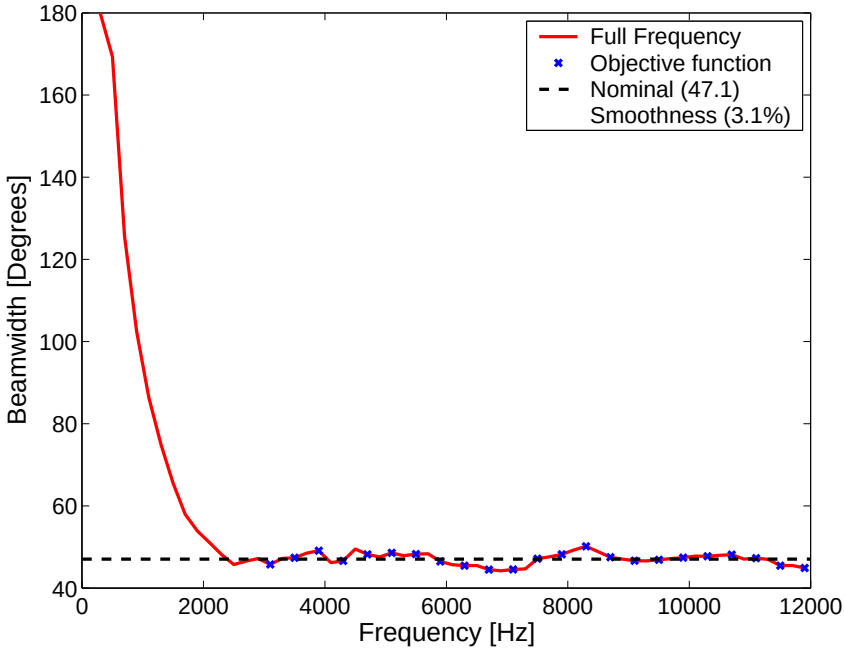
Figure 7.30 (b) respectively. This very small difference in geometry is probably smaller then the manufacturing tolerances of production horns, and the tiny improvement in the predictions of the numerical model probably do not warrant the extra computation effort required.

7.5.6 Conclusions

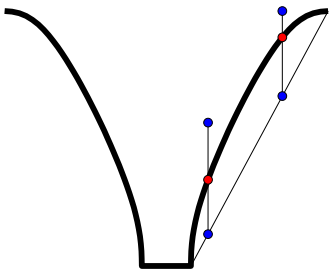
In summary, for a fixed set of general parameters (such as throat radius and horn length), it is possible to find simple 2 and 3 parameter horn geometries that are able to produce nearly constant beamwidth behaviour, at least over the frequency range studied. Efficient optimisation of parameter *S*, with approximately 50 objective function evaluations, is per-

formed with the Expected Improvement (EI) Infill Sampling Criteria (ISC) and the EGO optimisation algorithm, and the results are superior to the unconstrained optimisation performed in previous sections.

With the fixed parameters and unconstrained optimisation used in this section, the beamwidths found were unable to satisfy the nominal beamwidth constraint. This is most probably because the horn length and throat dimension, factors that control the overall angle of the horn, were fixed, and further investigation is required into ways to achieve constant beamwidth behaviour at the desired nominal beamwidth.



(a) Beamwidth



(x(1)=0.49,x(2)=0.69)

(b) Horn profile

Figure 7.28: Results of the unconstrained EI optimisation of Φ_3 for the 2 parameter complex spline horn geometry with a 2 inch throat and horn length 210 *mm.*

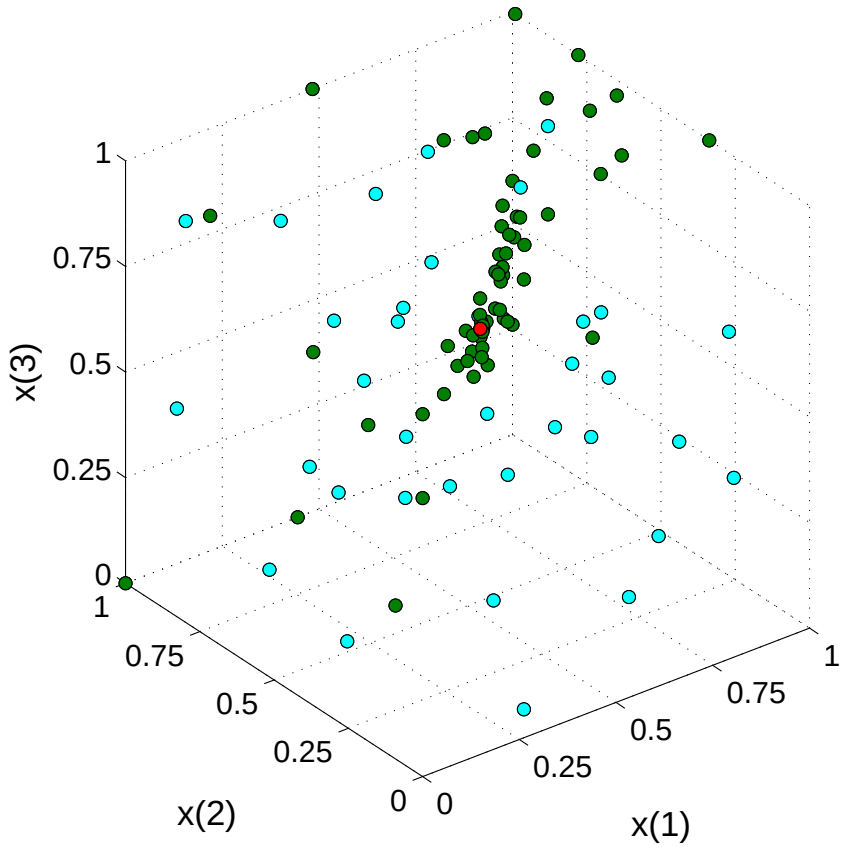
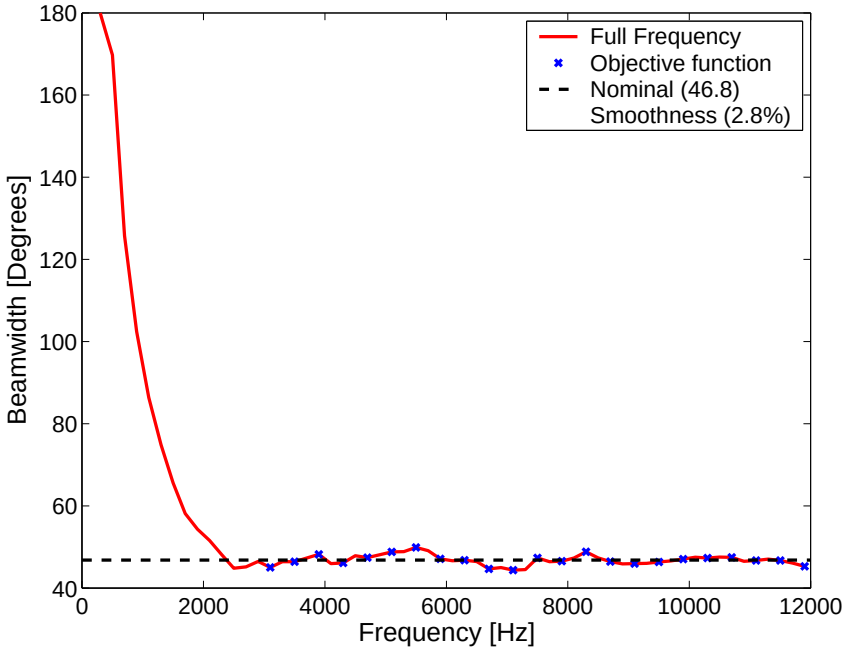
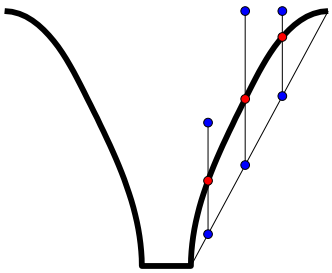


Figure 7.29: Optimisation trajectory for the 3 parameter complex spline horn geometry with a 2 inch throat and $L = 260$ mm. The cyan dots show the initial samples, green dots the EI sampling and the red dot shows the unconstrained global minimum.



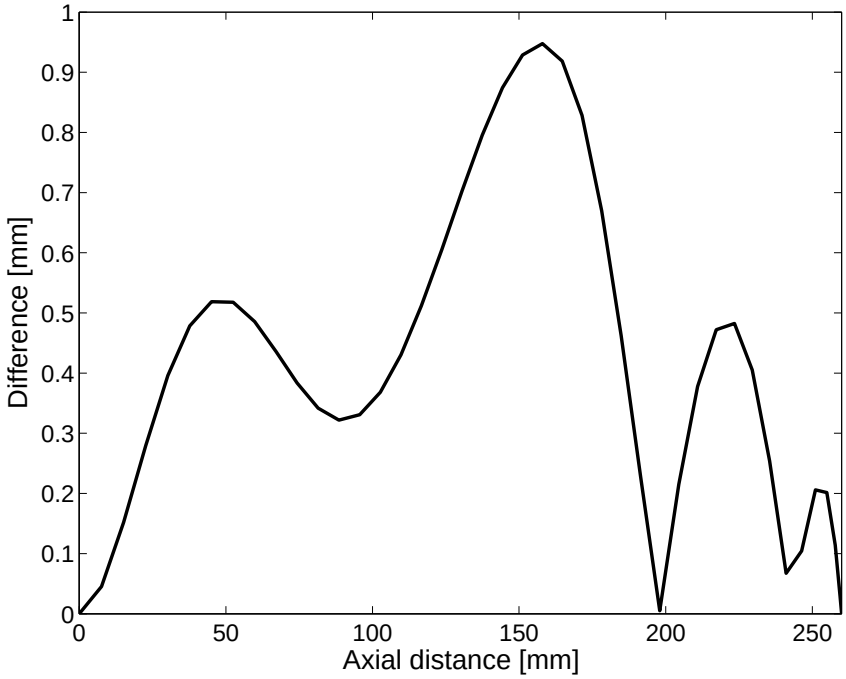
(a) Beamwidth



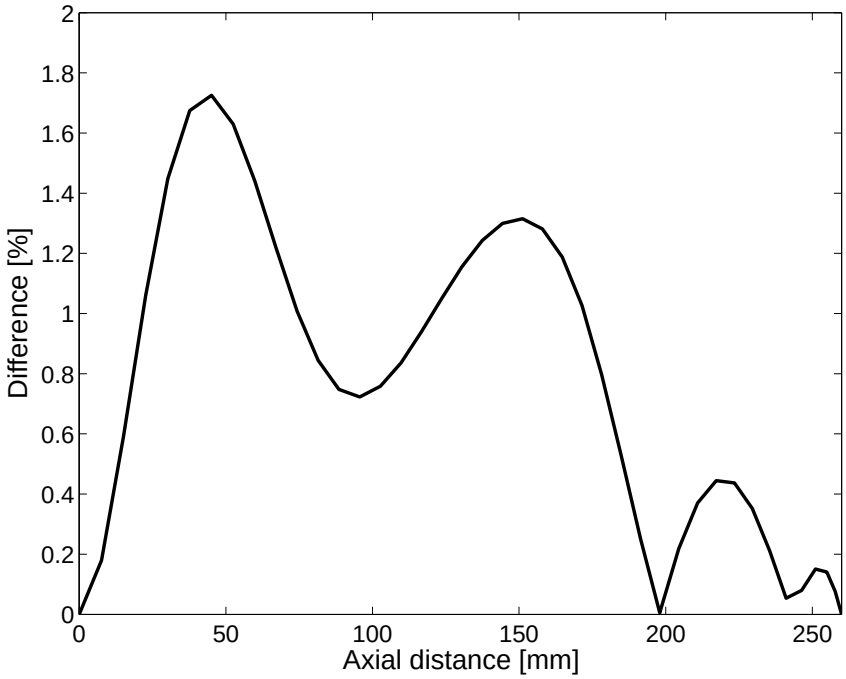
(x(1)=0.48,x(2)=0.43,x(3)=0.70)

(b) Horn profile

Figure 7.30: Results of the unconstrained EI optimisation of Φ_3 for the 3 parameter complex spline horn geometry with a 2 inch throat and horn length 210 *mm.*



(a) Absolute difference



(b) Percentage difference

Figure 7.31: Difference between the 2 parameter and 3 parameter radial profiles shown in Figure 7.28 (b) and Figure 7.30 (b) respectively.

7.6 Constant beamwidth horns

The work described in this section considers the design of a range of constant beamwidth horns. The work described in section 7.5.5 found that for a fixed throat radius and horn length, a 2 parameter Bézier spline based horn geometry was able to produce approximately constant beamwidth behaviour using EGO with an unconstrained EI ISC but the nominal beamwidth was not achieved. Unlike Section 7.3, with the specification of a design angle, it is not clear how the nominal beamwidth produced by this optimisation is controlled by the throat radius and horn length.

The technique described in Section 7.5.3 uses the constrained EGO approach to find the smoothest beamwidth possible for the geometry by varying 4 parameters including the throat radius and horn length. As can be seen from the beamwidth shown in Figure 7.24, the solution found is not particularly smooth. It is not known whether this is limited by the geometry (likely because one parameter is at the minimum) or by the optimisation procedure. Using a large multi-dimensional optimisation also limits possible insight into the visualisation and nature of the optimisation landscape.

One possible approach to this problem would be to embed an optimisation that found the smoothest possible geometry for a given throat radius and horn length as an objective function for a second optimisation. This optimisation would then vary the throat radius and horn length until it found the desired nominal beamwidth. An unconstrained EGO optimisation would be able to perform the optimisation quite simply, provided that the desired nominal beamwidth is physically realisable for the given ranges of throat radii and horn lengths. The cost of a single objective function evaluation for the second optimisation is extremely large (in the order of 1 hour for a Pentium 4 2.8 GHz running Windows XP). If an optimisation was performed using this approach, with up to 50 objective function evaluations would be used to find a single solution. If different value of nominal beamwidth was required, then another expensive optimisation must be performed.

The solution adopted in this thesis does not perform a second optimisation. Instead the results of a series of constant beamwidth optimisations at a number of different throat radii and horn length parameters are interpolated using the same technique as used in the EGO optimisation method (Kriging). The entire range of throat radii and horn length parameters is sampled, with parameters chosen using space filling sampling techniques (Appendix B.5). This solution is possible because the dimension is small (2 parameters), and the Kriging interpolator is very efficient. This solution effectively produces a "design chart" for axisymmetric constant beamwidth horns where the horn designer can "look up" designs and make trade offs between horn dimensions and nominal beamwidths.

The upper and lower bounds of the variables, the mouth to throat radius ratio, $\frac{R_m}{R_t}$, and the horn length, L, are given in Table 7.4. A set of 50 parameters between the upper and lower bounds has been selected with using the space sampling Improved Hypercube Sampling (IHS) method. At each of these parameters, a second optimisation to find the smoothest possible beamwidth with the throat radius and horn length held constant is performed. Kriging models were fit to the results of the second optimisation (parameter S, nominal beamwidth Φ_2, as well as the values of the constant beamwidth optimisation, $x(1)$ and $x(2)$) and the mean square error of the resulting approximations examined by hand to find appropriate places to sample next, because the dimension of the current problem is small, and visualisation is easy. For more complicated problems, or for a more automatic way of finding new places to sample, the MAXVAR ISC could be used. A number of new sample points was chosen, and further constant beamwidth optimisations performed, to reduce the error in the interpolation. Overall 81 constant beamwidth optimisations were performed.

Variable	Description	Lower	Upper
$\frac{R_m}{R_t}$	Mouth to throat ratio	2	13
L	Horn length	200 mm	400 mm

Table 7.4: Upper and lower bounds of parameters used to describe the constant beamwidth horn geometry.

Results for the parameter S are shown in Figure 7.32 as a contour plot. Low values of the parameter S means that the standard deviation of the beamwidth normalised by the mean value of the beamwidth is low, and close to constant beamwidth behaviour is achieved. The 50 cyan dots represent the initial sample positions, and the 31 green dots represent additional samples chosen to improve the Kriging interpolation. The value of the parameter S is low, across most of the plot, except for low values of $\frac{R_m}{R_t}$. The reason for these large values of S, and non-constant beamwidth behaviour, is evident when examining typical geometries produced for low values of $\frac{R_m}{R_t}$ such as those shown in Figure 7.33. The smoothest profile found by the optimisation routine for the low values of $\frac{R_m}{R_t}$ is one where the value of the parameter that controls the first spline control point, $x(1)$, is small, and the horn profile has a point of inflection near the throat. To suppress this anomalous behaviour, only the values of S below 4 are to be considered, the approximate boundary between the red (bad) and blue (good) areas on Figure 7.32. It is suggested that future studies examine different objective functions to eliminate this behaviour (Section 8.3).

Figure 7.34 shows contours of parameter S for varying throat radius and horn length, with the black mask covering the anomalous horn shapes. This plot, along with Figure 7.35, a contour of the nominal beamwidth, can be considered the main results of this thesis. They provide a method to design a horn that is as smooth as possible for a given nominal beamwidth. Figure 7.36 overlays a contour line plot of nominal beamwidth over a plot of smoothness. Picking a particular contour of nominal beamwidth and moving along it until the minimum smoothness is chosen gives the values of $\frac{R_m}{R_t}$ and L. These design charts have been generated by a Kriging interpolator, a computer model that can easily be interfaced with optimisation techniques to find a minimum solution automatically. However, visualisation of the search space allows the designer to see many potential solutions and pick the best one. These can then be used with the plots in Figure 7.37 to find the optimal value of $x(1)$ and $x(2)$. These values can then be used to construct the horn profile. No expensive optimisation using the EGO and the source superposition technique is needed.

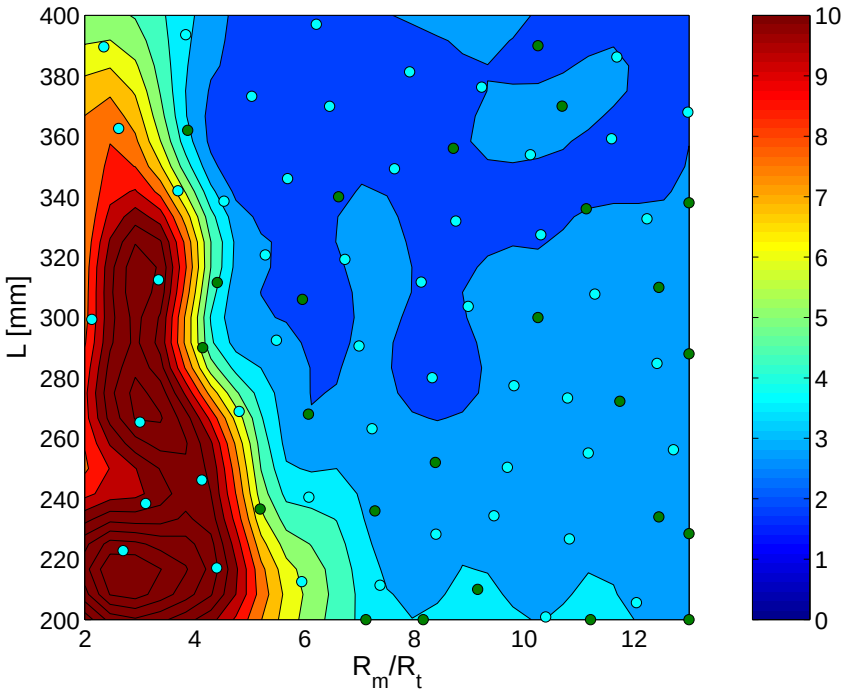


Figure 7.32: Contour of parameter S. The cyan dots show the initial samples and the green dots additional sampling to reduce uncertainty in the interpolation. The red contours show anomalous behaviour and should be suppressed.

The MATLAB DACE Kriging interpolation system (Lophaven et al., 2002b) is able to calculate vector interpolations, and the beamwidth of the system can also be stored. To check the validity of the optimum interpolation system, a design for a 60° beamwidth horn is undertaken. Figure 7.36 is used to estimate the performance of a 60° horn. A point $\frac{R_m}{R_t} = 12.45$ and $L = 220$ mm is selected as a good design. The Kriging approximation then gives $x(1) = 0.43$ and $x(2) = 0.68$ as predictions of optimal solutions. The horn profile calculated using these values is shown in Figure 7.38 (b). The true solution is calculated using this profile and shown in Figure 7.38 (a), plotted with non-dimensional frequency kR_m (where $k = 2\pi f/c$), along with the Kriging prediction of the beamwidth. As can be seen, the Kriging approximation to the beamwidth and the true beamwidth are very similar. The performance of the horn is good enough for design purposes, and because all parameters are scaled by the horn mouth dimension R_m, the results can be easily scaled to any frequency range.

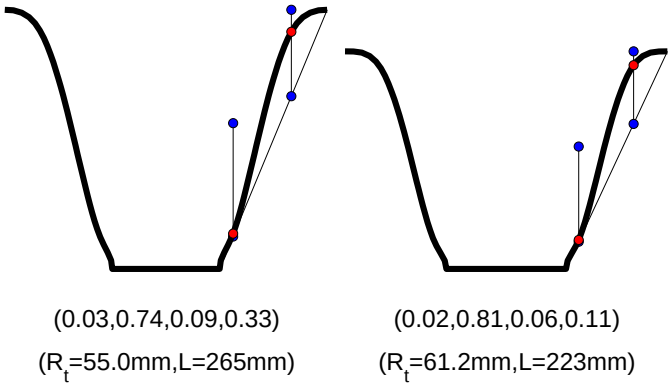


Figure 7.33: Typical optimum horn profile for small values of $\frac{R_m}{R_t}$.

Another use of the design tools is presented. Suppose a fixed length of horn is required of $L = 260$ mm. By fixing this dimension, then picking a good dimension for $\frac{R_m}{R_t}$ from Figure 7.36 (in this case $\frac{R_m}{R_t} = 10.8$, giving $x(1) = 0.50$ and $x(2) = 0.69$), an optimal solution is found. This is shown in Figure 7.39, a good design for a horn restricted to $L = 260$ mm.

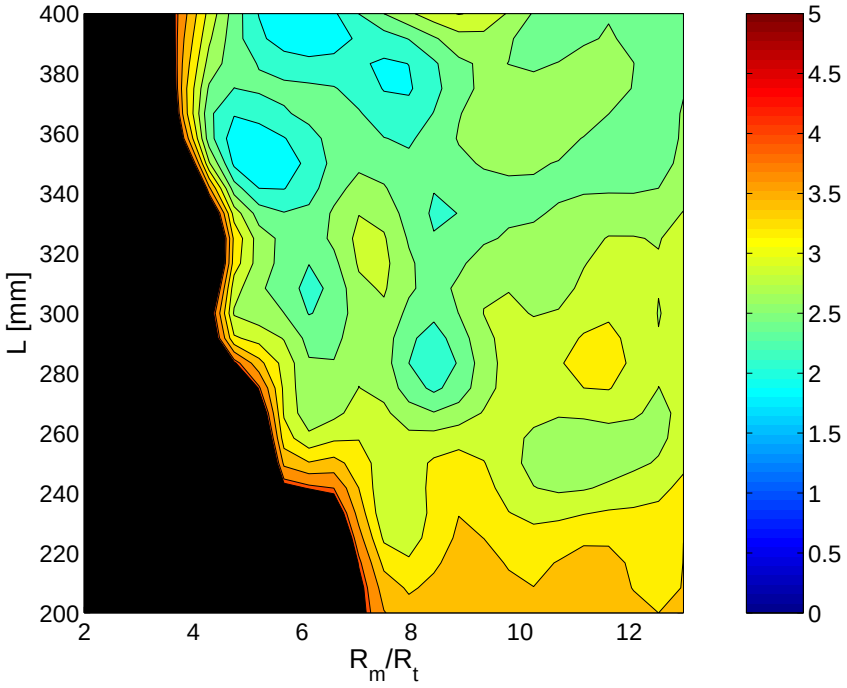


Figure 7.34: Contour of the parameter S, where small values imply better "constant beamwidth behaviour". The black contour masks anomalous behaviour.

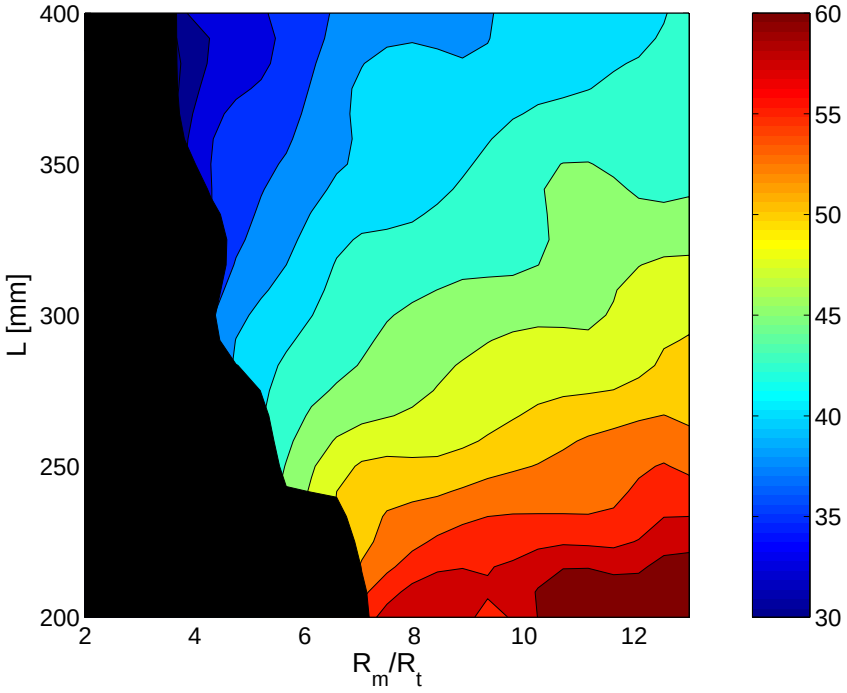


Figure 7.35: Contour of the nominal beamwidth Φ_2, showing the range of constant beamwidth horns achieved. The black contour masks anomalous behaviour.

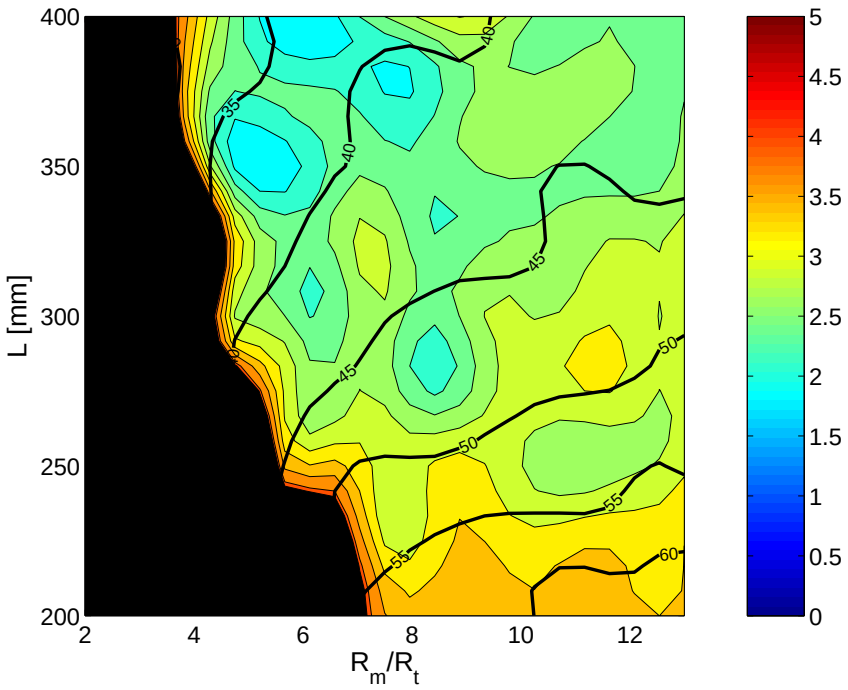
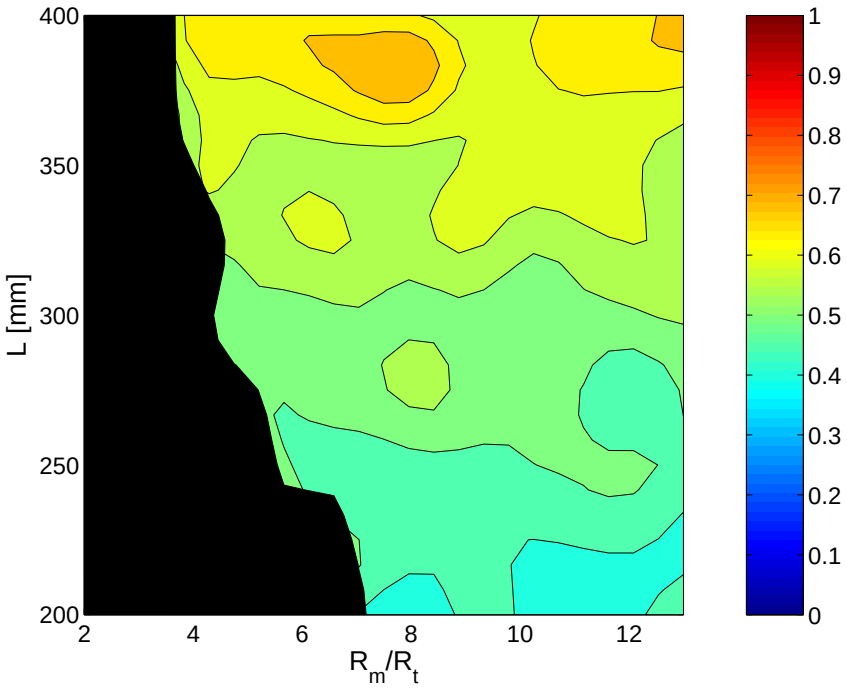
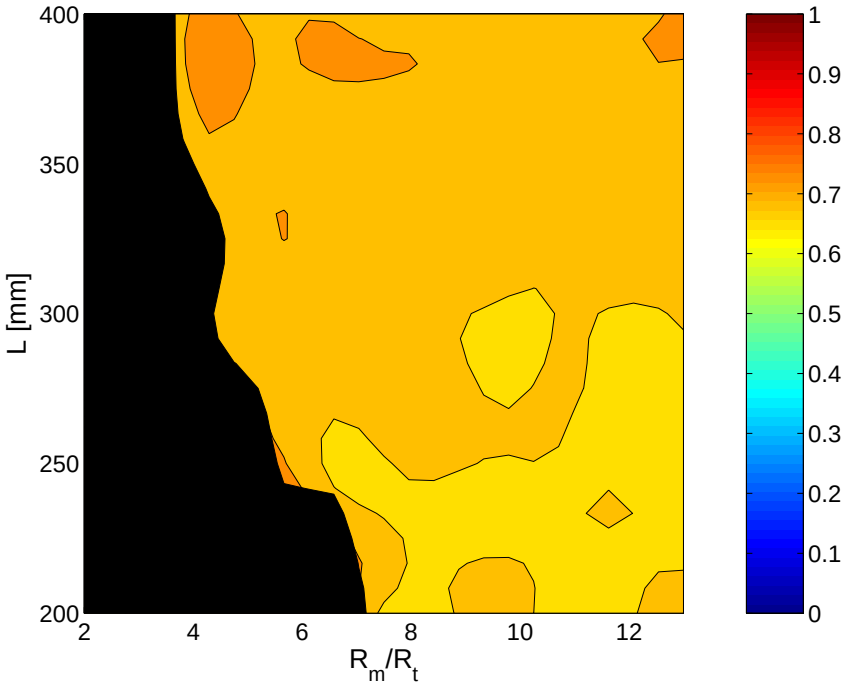


Figure 7.36: Constant beamwidth horn "design chart". A contour of the parameter S is overlaid with contour of the nominal beamwidth Φ_2. The black contour masks anomalous behaviour.

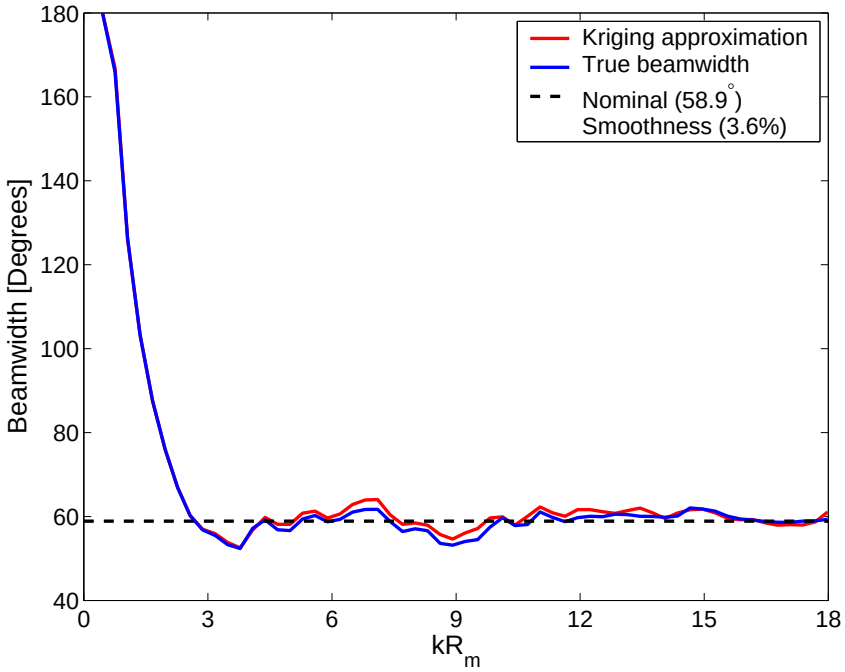


(a) $x(1)$

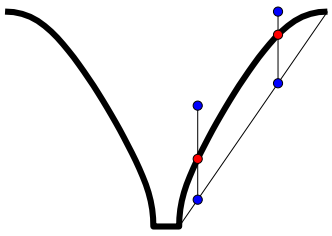


(b) $x(2)$

Figure 7.37: Contours of the parameters $x(1)$ and $x(2)$ that define the shape of the constant beam-width horns.



(a) Beamwidth

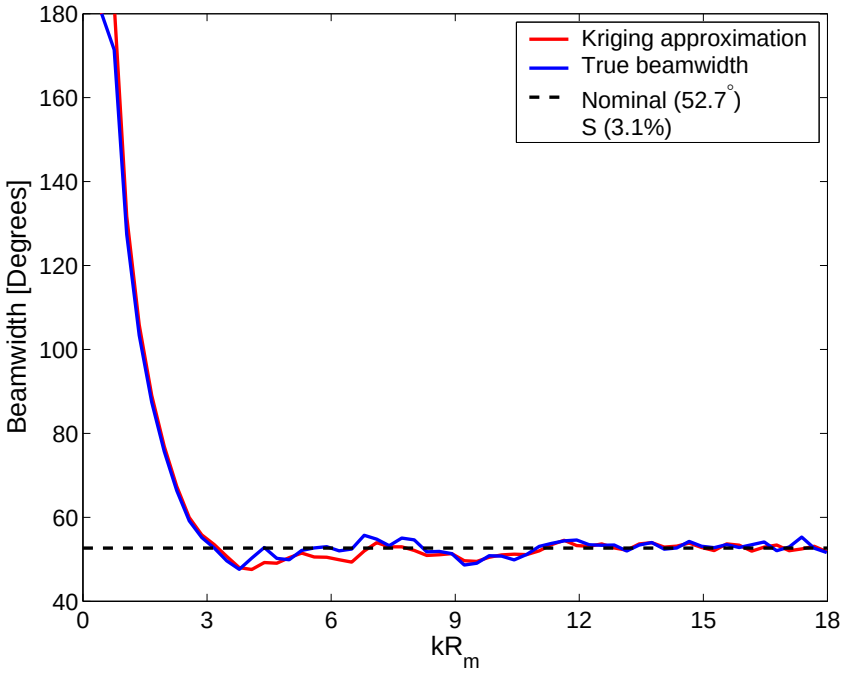


(0.43,0.68,0.95,0.10)

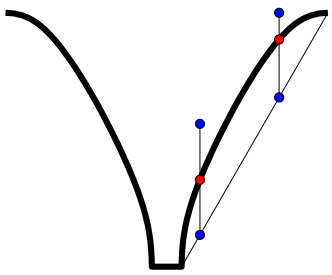
(R_t=13.3mm,L=220mm)

(b) Horn profile

Figure 7.38: Results of the Kriging approximation compared to a calculation for a 60° *constant beamwidth horn calculated using the design chart.*



(a) Beamwidth



(0.50,0.69,0.80,0.30)

(R_t=15.3mm,L=260mm)

(b) Horn profile

Figure 7.39: Results of the Kriging approximation compared to a calculation for a constant beamwidth horn of length $L = 260$ mm calculated using the design chart.

7.7 Conclusions

Optimisation techniques have been developed that can find the smoothest possible beamwidth response as a function of frequency for a given horn parameterisation, and often this behaviour can be considered to approach "constant beamwidth behaviour". If the horns are exhibiting true constant beamwidth behaviour then the frequency response of the horn will be smooth (Rogers and Van Buren, 1978). This remains to be tested both numerically and experimentally (Section 8.3).

If the geometry is physically able to create the desired nominal beamwidth, then optimisation techniques are available that will find the smoothest possible beamwidth. One approach is to use the constrained EGO technique to find a single point in space that both satisfies the beamwidth constraint and is as smooth as possible. This technique, while it does find an optimum solution, is time consuming and tells nothing about how the parameters vary. A more satisfying approach is to calculate the smoothest possible beamwidth for a range of horn throat dimensions and lengths, and then fit a Kriging interpolator to find the values in between. This way, the information calculated can be reused, and new designs calculated without further expensive optimisation.

There is much more that can be done with the optimisation of horn loaded loudspeakers:

- Objective functions that better capture constant beamwidth behaviour would provide more robust optimisation. One possible approach would be to fit a function that describes constant beamwidth behaviour with a number of parameters, such as nominal beamwidth and low frequency beamwidth performance. A least squares fit to the beamwidth data would give the parameters, and the correlation coefficient (Weisstien, 2004) would be a non-dimensional measure of how close constant beamwidth behaviour was achieved.
- More general horn geometry parameterisations may allow smoother horn beamwidths.

- The horns investigated had uniform velocity distributions applied at the throat, however this is known not to be the case above a certain frequency (Behler and Makarski, 2003). The effect of real compression drivers should be investigated.
- It should be possible to tailor the compression driver response to help achieve constant beamwidth behaviour (Geddes, 2002) or to design horn loaded tweeters.
- Horns with different beamwidths for different axes (3D horns) would be more industrially relevant.

However, these problems are beyond the scope of the current study (see Section 8.3).

Chapter 8

Conclusions and recommendations

This chapter reiterates the aims given in the first chapter and summarises the work completed in this thesis. It clearly states the contributions to current knowledge in the optimisation of horn loaded loudspeakers, and gives recommendations for future work.

8.1 Introduction

The introductory chapter describes the aims of horn design for cinema loudspeakers: to produce an easily specified frequency independent beamwidth, and to provide a smooth frequency response over as large a bandwidth as possible. The overall aim of this thesis was to develop fast and reliable optimisation techniques for horn loaded loudspeakers to achieve a good horn design method for cinema loudspeakers.

This aim was achieved by first examining the literature to find out what techniques had been used previously. It was found that there are no analytical or semi-analytical techniques suitable for the design of horns for cinema loudspeakers, that fast numerical techniques are necessary and that the numerical method chosen must be able to include the

effects of higher order modes propagating within the horn. It was also found that although studies had previously attempted to optimise horn loaded loudspeakers, none of these techniques were suitable for cinema loudspeaker systems.

The specific aims of this thesis were to;

- **Examine experimentally the nature of the sound field at the mouth of representative axisymmetric horns (near field) and the horn beamwidth (far field).**
  The experimental results for the near field sound pressure generated by a horn found that above a certain limiting frequency, plane waves ceased to exist at the mouth of the horn, and the sound field is quite complex. This implies that any numerical method used must be able to model this complex field implicitly. Far field experimental results have been used to validate the numerical methods. This aim has been achieved.

- **Develop fast and accurate numerical models of horn loaded loudspeakers.**
  Of the many numerical techniques available, the source superposition technique of Koopmann and Fahnline (1997) is a good choice for modelling the sound field radiated by horns. Work described in this thesis demonstrated it to be significantly faster than traditional BEM techniques and also suitable for modelling thin structures such as horns. The technique is capable of reproducing the sound field generated by a horn loaded loudspeaker from a specification of the horn geometry, and the accuracy of the reproduction is adequate for design purposes within the specified frequency range. This aim has been achieved.

- **Develop fast and reliable optimisation techniques for horn loaded loudspeakers.**
  A theory used to design a Constant Beamwidth Transducer (CBT) for sonar applications has been examined in the context of horn design. This has shown that a frequency independent beamwidth is physically realisable. The theory also states

that the frequency response of such a transducer, once it has achieved a frequency independent beamwidth, will be smooth. This implies that a horn that has achieved a frequency independent beamwidth in the same manner at the CBT transducer will also have a smooth frequency response, and that the major aim of horn design should be to produce a frequency independent beamwidth.

Fast and reliable gradient free optimisation techniques for expensive objective function evaluation have been developed, along with objective functions that quantify the aims of a horn designer. These have been applied to horn loaded loudspeakers and horn geometry based on a 2 parameter Bézier spline has been developed. The optimisation techniques have been used to develop a design method for horn loaded loudspeakers. The optimum beamwidth for a wide range of sizes have been pre-calculated and a series of design graphs created, allowing optimal designs to be quickly and easily created. This aim has been achieved.

8.2 Contributions to current knowledge

The work in this thesis makes the following contribution to the state of current knowledge in acoustic horn theory:

- Foremost from an industrial perspective, the work described in this thesis provides an optimisation method for the design of constant beamwidth horn loaded loudspeakers. This provides smooth control of the beamwidth over a range of frequencies governed by the size of the horn. A design method that allows an optimal design to be chosen quickly and easily from a series of graphs or a simple computer programme has been developed.

- Fast numerical methods have been developed and validated for analysing horn geometries. This has been done using the source superposition method, a numerical

technique that is similar to traditional BEM but is significantly faster. Comparisons have been made with experimentally measured sound pressure data in the far field and other numerically based sound pressure predictions, such as boundary element techniques, to find the limits of validity for this approach. An extensive literature search failed to find any evidence that this technique has previously been applied to the numerical modelling of horns.

- Constant Beamwidth Transducer (CBT) theory has been reviewed in the context of acoustic horns, and an optimum velocity profile developed using an optimisation technique. This shows that the aim of a frequency independent beamwidth is physically realisable under some circumstances.

- The nature of the near field of the horn has been examined. The existence of higher order modes above a limiting frequency has been demonstrated, which has implications for the technique chosen to model the horn.

The work in this thesis makes the following contribution to the state of current knowledge in numerical methods in acoustics:

- For both the source superposition technique and the traditional BEM, it has been found that accurate far field sound pressure results can be obtained with a much reduced mesh density than has been reported previously in the literature, at least for the cases considered here. This is a significant finding, as the efficiency of both techniques decreases rapidly with increasing model size.

- The source superposition technique has been found to generate matrices that are extremely diagonally dominant, and hence are well suited to solution with an iterative technique such as GMRES. Traditional BEM has had to use a preconditioning technique (Migeot et al., 2000, Chen, 1999) to effectively use iterative solution techniques. The advantage of this finding will become more apparent as the problem

size becomes larger. For the problems considered in this thesis, element assembly time is as dominant as solution time.

The work in this thesis makes the following contribution to the state of current knowledge in global optimisation methods:

- The Regional Extreme (RE) Infill Sampling Criteria (ISC) of Sasena (2002) has been compared with the criteria originally proposed by Watson and Barnes (1995). Sasena's interpretation was found to be different to that of the original implementation of the Regional Extreme criteria. The original interpretation is found to be exactly equivalent to Schonlau's (1997) Expected Improvement. It was also found here that an alternative implementation of Sasena's criteria (which cannot be called a regional extreme criteria) may not find the global optimum.

8.3 Recommendations for future work

From an industrial perspective, there are a number of tasks that would be very useful:

- The most pressing need would be the construction and measurement of horn designs given in Chapter 7 to experimentally verify constant beamwidth behaviour.
- A procedure directed at the design of horns that allow independent control of the beamwidth in different axes. This is not possible with an axisymmetric geometric specification described in Chapter 7. However an extension of this geometry to a second axis should be relatively simple. CBT theory has been shown to be extend in a simple manner to the second axis (Keele, 2000), so independent axis control is physically realistic. A question remains as to the efficiency of the optimisation technique, as the number of variables to be optimised doubles.

- The development of a simple computer program that embodies the "design chart" given in Section 7.6 would be very useful for the design of axisymmetric horn loaded loudspeakers.

- Modifications to the boundary conditions at the throat of the horn, by changing the design of the compression driver, may help achieve constant beamwidth behaviour. This was first suggested by Geddes (2002). The design of constant beamwidth horn loaded tweeters is a related problem.

From an academic perspective, much work can be done with the optimisation technique.

- The objective function can be improved to better represent constant beamwidth behaviour. By fitting a suitable parameterised curve through the calculated beamwidth points, the correlation coefficient (Weisstien, 2004) can be used as a scale free measure of the smoothness of the beamwidth. This would provide a more robust optimisation.

- The constrained EGO algorithm is not robust. The method developed here for constrained optimisation works well for the applications described in this thesis, as the MAXVAR sampling of the constrained area is generally restricted to a small subset of the whole parameter space. A more general Infill Sampling Criteria (ISC) such as Schonlau's (1997) Expected Improvement (EI) is not amenable for constrained optimisation with either DIRECT or Adaptive Simulated Annealing. Research into optimisation methods able to perform constrained optimisation of the Expected Improvement ISC would be most valuable.

- Alternatively, as the equality constrained SQP algorithm has been found to be very efficient when robust gradient information is available, further research on an efficient method to calculate the gradient of the source superposition solution would be very valuable.

Modelling horn loaded loudspeakers can be improved in a number of ways.

- A simple modification to the acoustic power calculation algorithm in the source superposition code to include 1/4 symmetry planes would allow the rapid calculation of the power response of horn loaded loudspeakers. The frequency response of a constant beamwidth horn could then be investigated.
- The choice of boundary conditions can be questioned. It has been found that above a certain frequency, plane waves do not exist at the throat of the horn (Behler and Makarski, 2003). The plane wave boundary condition did not effect the results of the conical horn model when compared to experimental beamwidth results, possibly due to the sharp interface at the step generating higher order modes. However the exponential horn shows a discrepancy above a certain frequency. The inclusion of a more accurate model of the compression driver should remove this discrepancy.

The source superposition technique is an efficient technique for predicting the power output from radiating structures, and in this thesis it has been found to be an excellent tool for modelling the far field pressure distribution of horn loaded loudspeakers. There is much potential for the technique to be extended and investigation into similar boundary element like techniques may result in better numerical methods.

- The source superposition technique can be extended immediately to predicting scattering from rigid structures (see Ochmann 1999, Section 1) and the technique in combination with the GMRES iterative solver would be very efficient. Further work would be necessary to extend the technique to general impedance boundaries.
- To further increase the efficiency of the source superposition technique for multiple frequency calculations, a matrix interpolation technique such as described in Section (4.4.4) should be investigated further, especially in light of the Kriging (Appendix B.4) interpolation method, which is a fast and efficient interpolation technique used as a component of the global optimisation method developed here.

- The choice of linear elements in the source superposition technique is also an issue in accurately resolving the surface area at low element densities (Koopmann and Fahnline, 1997). The use of quadratic elements and shape functions would eliminate this problem.

- An "element agglomeration" technique (Fahnline, 1995) can be used to reduce the size of the matrix generated by the source superposition technique, especially at low frequencies. This process does not eliminate the expensive integration of the singular functions on the surface, but does reduce the matrix size. A problem with this is the automatic selection of these larger agglomerated super-elements. Techniques such as those used in multigrid solvers (Moulitsas and Karypis, 2001) and graph partitioning (Karypis and Kumar, 1998) may go some of the way to providing an automatic solution.