Article

Binaural Range Finding from Synthetic Aperture Computation as the Head is Turned

Duncan Tamsett

Environmental Research Institute, North Highland College, University of the Highlands and Islands, Thurso, Caithness KW14 7EE, UK; duncan.tamsett@uhi.ac.uk; Tel.: +44-(0)-1847-889-589

Abstract: A solution to binaural direction finding described in Tamsett (*Robotics* **2017**, 6(1), 3) is a synthetic aperture computation (SAC) performed as the head is turned while listening to a sound. A far-range approximation in that paper is relaxed in this one and the method extended for SAC as a function of range for estimating range to an acoustic source. An instantaneous angle λ (lambda) between the auditory axis and direction to an acoustic source locates the source on a small circle of colatitude (lambda circle) of a sphere symmetric about the auditory axis. As the head is turned, data over successive instantaneous lambda circles are integrated in a virtual field of audition from which the direction to an acoustic source can be inferred. Multiple sets of lambda circles generated as a function of range yield an optimal range at which the circles intersect to best focus at a point in a virtual three-dimensional field of audition, providing an estimate of range. A proof of concept is demonstrated using simulated experimental data. The method enables a binaural robot to estimate not only direction but also range to an acoustic source from sufficiently accurate measurements of arrival time/level differences at the antennae.

Keywords: robotic sensing; acoustic localization; binaural systems; synthetic aperture

---

1. Introduction

An aspect of importance in some classes of robots is an awareness and perception of the acoustic environment in which a robot is immersed. A prerequisite to the application of higher level processes acting on acoustic signals such as acoustic source classification and speech recognition/interpretation is a process to locate acoustic sources in space [1–23]. A method for a binaural robotic system to locate directions to acoustic sources, based on a synthetic aperture computation, is described in Tamsett [24]. A far-range approximation in that paper is relaxed in this one and the method is generalized for potentially determining range to an acoustic source as well as its direction.

With two ears, humans extract information on the direction to acoustic sources over a spherical field of audition based on differences in arrival times of sound at the ears (interaural time difference or ITD) for frequencies less than approximately 1500 Hz [25–27]. Measurement of arrival time difference might be made by applying a short time-base cross-correlation process to sounds received at the ears [28] or by a functionally equivalent process [29–35]. For pure tones at higher frequencies locating acoustic sources is dominated by the use of interaural level difference (ILD) [25–27].

Other aural information is also integrated into an interpretation of the direction to an acoustic source. The spectral content of a signal arriving at the ears is affected by the shape of the pinnae and the head around which sound has to diffract—the so-called head related transfer function (HRTF) [36–39]—and this effect provides information that can be exploited for aural direction finding.

Measurement of an instantaneous time or level difference between acoustic signals arriving at a pair of listening antennae or ears allows an estimate of the angle λ between the auditory axis and

the direction to an acoustic source. This ambiguously locates the acoustic source on the surface of a cone with an apex at the auditory center and an axis shared with the auditory axis. The cone projects onto a circle of colatitude of a spherical shell sharing a center and an axis with the auditory axis (a lambda circle). Based on observations of human behavior, Wallach [40] inferred that humans disambiguate the direction to an acoustic source by "dynamically" integrating information received at the ears as the head is turned while listening to a sound (also [41]). A solution to the binaural location of the directions to acoustic sources in both azimuth and elevation based on a synthetic aperture computation is described in Tamsett [24] representing the "dynamic" process posited by Wallach [40].

The synthetic aperture computation (SAC) approach to finding directions to acoustic sources applied to a pair listening antennae as the head is turned provides an elegant solution to the location of the direction to acoustic sources. This process, or in nature a neural implementation equivalent to it, is analogous to those performed in anthropic synthetic aperture radar, sonar, and seismic technologies [42], and in the same way that SAC in these technologies considerably improves the resolution of targets in processed images over that in unprocessed "raw" images, so the ambiguity inherent in locating an acoustic source with just two omnidirectional acoustic antennae, collapses from a multiplicity of points on a circle for a stationary head, to a single point along a line, by deploying a SAC process as the head is turned while listening to a sound.

Only relatively recently has binaural sensing in robotic systems, in emulation of binaural sensing in animals, developed sufficiently for the deployment of processes for estimating directions to acoustic sources [1–13]. Acoustic source localization has been largely restricted to estimating azimuth [1–13] on the assumption of zero elevation, except where audition has been fused with vision for estimates also of elevation [14,15,17,18]. Information gathered as the head is turned has been exploited either to locate the azimuth at which the ITD reduces to zero, thereby determining the azimuthal direction to a source, or to resolve the front–back ambiguity associated with estimating only the azimuth [4–14,19,20]. Recently, Kalman filters acting on a changing ITD have also been applied in robotic systems for acoustic localization [21–23].

The second aspect to locating acoustic sources is estimating range. Some species of animal possess ears with pinnae having a response to sound that is strongly directional [43], and that are independently orientable, for example cats. Such animals can explore sounds by rotating the pinnae while keeping their heads still. In this way they are able to determine the direction to an acoustic source with a single ear and in principle using both ears estimate range by triangulation.

Animals without this facility (e.g., humans and owls) must use other mechanisms or cues for range finding. Owls are known to be able to catch prey in total darkness [44] suggesting they are equipped to accurately measure distance to an acoustic source. An owl flying past a source of sound could use multiple determinations of direction to estimate range by triangulation. Or in flying directly towards a point acoustic source an owl will experience a change in intensity due to an inverse square relationship with range which could be exploited for monaural estimates of range. For example, a quadrupling in intensity during a swoop on a point acoustic source would indicate that the distance to the source has halved; and in this way an estimate of instantaneous current range could be made. In addition, sounds that are familiar to an animal have an expected intensity and spectral distribution as a function of range and this learned information can be exploited for estimating range [45,46]. Range in robots has been estimated on the basis of triangulation to acoustic source involving lateral movement of the head [6].

A far-range approximation in Tamsett [24] restricts acoustic source localization to direction. It was suggested [24] that SAC in principle could provide an estimate of range from near-field deviations from far-field expectations. In relaxing the far-range approximation in the current paper, a synthetic aperture computation as a function of range is formulated to enable range to acoustic source as well as direction to be estimated.

A proof of concept demonstrating the principle and potential utility of the method is provided through the use of simulated experimental data. Multiple SACs as a function of range are performed

on the simulated data and distance to acoustic source estimated by optimizing range for the set of lambda circles of colatitude generated as a function of range that intersect to best converge or focus to a point. Employing a SAC process in this way adds a dimension to the SAC process that finds only direction to source. Acoustic energy maxima are sought effectively in a three-dimensional virtual acoustic volume rather than over a two-dimensional acoustic surface.

The solution could be implemented in a binaural robotic system capable of generating a series of sufficiently accurate values of λ estimated from measurements of arrival time/level differences as the head is turned. Any implementation in nature will involve biological/neural components to provide an equivalent to the mathematics-based functions deployed in an anthropic robotic system.

2. Arrival Time Difference, Angle to Acoustic Source, and Range

A straight line simplification of the relationship between arrival time difference, angle to source, and range is illustrated in Figure 1. A more elaborate model might allow for diffraction around the head to reach the more distant ear [19,47].
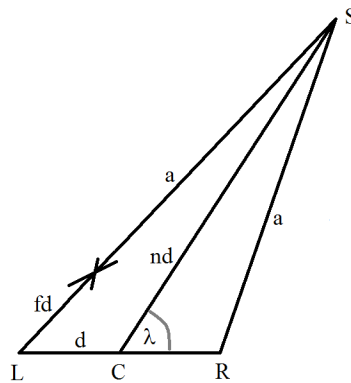
Figure 1. Top view on an auditory system with a left (L) and right (R) ear receiving incoming horizontal acoustic rays from a source (S) a finite distance from the auditory center (C).

In Figure 1:

L   represents the position of the left ear, and R the right ear; the line LR lies on the auditory axis;
C   represents the position of the auditory center;
S   is the position of the acoustic source;
λ is the angle at the auditory center between the auditory axis and the direction to the acoustic source;
d represents the distance between the ears (the length of the line LR);
nd represents the distance of the acoustic source from the auditory center as a multiple n of the length d (the length of the line CS);
fd represents the difference in the acoustic ray path lengths from the source to the ears as a proportion of the length d ($-1 \leq f \leq 1$); and
a is the distance from the acoustic source to the right ear (the length of the line SR).

The distance fd is related to the difference in arrival times at the ears measured by the auditory system by:

$$fd = c\,\Delta t \tag{1}$$

where:

c is the acoustic transmission velocity (e.g., 330 ms^{-1} for air); and
Δt is the difference in the arrival time of sound received at the ears.

Applying the cosine rule to the triangle SCR:

$$a^2 = (d/2)^2 + (nd)^2 - 2\,(d/2)\,(nd)\,\cos\,\lambda \tag{2}$$

$$a^2 = d^2/4 + n^2 d^2 - nd^2\,\cos\,\lambda \tag{3}$$

Applying the cosine rule to the triangle SLC:

$$(a + fd)^2 = (d/2)^2 + (nd)^2 - 2\,(d/2)\,(nd)\,\cos\,(\pi - \lambda) \tag{4}$$

$$a^2 + f^2 d^2 + 2afd = d^2/4 + n^2 d^2 - nd^2\,\cos\,\lambda \tag{5}$$

Substituting for *a* in Equation (5), using Equation (3) yields:

$$f^2 + f\,2\,(1/4 + n^2 - n\,\cos\,\lambda)^{1/2} - 2\,n\,\cos\,\lambda = 0 \tag{6}$$

For infinite range (*n* = ∞) Equation (6) reduces to:

$$f = \cos\,\lambda \tag{7}$$

Equation (7) is the relationship between *f* and *λ* for the far-range approximation [24] in which acoustic rays incident on the ears are parallel rather than diverging from a point a finite distance from the ears.

Equation (6) is quadratic in *f* yielding a single physically realizable solution:

$$f = (1/4 + n^2 + n\,\cos\,\lambda)^{1/2} - (1/4 + n^2 - n\,\cos\,\lambda)^{1/2} \tag{8}$$

Equation (8) may be rearranged for a solution to *λ* in terms of *n* and *f*:

$$\lambda = \mathrm{acos}(\,(f^2 + (f^2 - f^4)/4n^2\,)^{1/2}\,) \tag{9}$$

Equation (9) reduces to Equation (7) for infinite range as expected. A value for *λ* computed for values of *f* and *n*, defines a circle of colatitude on a spherical surface of radius *nd* sharing a center and axis with the auditory axis.

3. Synthetic Aperture Calculation Range Finding: Simulated Experimental Results

For an instantaneous estimate of *λ* as the head is turned, a measure of acoustic energy may be extended over the corresponding lambda circle and integrated into a virtual field of audition. In a SAC, as the head is turned through an angle $\Delta\theta$, the current state of the integration of acoustic energy over previous instantaneous lambda circles in the virtual field of audition, is rotated by $-\Delta\theta$, and the current measure of acoustic energy at the ear extended over the circle of colatitude for the current value of *λ*, and added into the field. If multiple sets of lambda circles are generated for multiple possible values for range *n*, an estimate of range is provided by the optimal value of range for which the integral of acoustic energy over the corresponding set of lambda circles yields a maxima; or alternately, in which the corresponding set of lambda circles best converges/focuses to a point where the circles intersect. This constitutes a SAC in a three-dimensional virtual field of audition.

3.1. Horizontal Auditory Axis

A proof of concept and the utility of the method for locating range to an acoustic source are demonstrated by performing SAC on simulated experimental data. Consider a source of sound at an inclination relative to the horizontal of $\varphi = -30°$, and a range of *n* = 3.5 times the distance between

the ears. Simulated values for λ corresponding to an azimuthal disposition of the head about a vertical axis relative to the longitudinal position of the source θ may be computed [24] from:

$$\lambda = \mathrm{acos}(\sin\theta\cos\varphi) \tag{10}$$

The value of θ is positive when the direction to the source is clockwise relative to the direction the head is facing. Corresponding simulated values for f are computed using Equation (8). Data for five values of θ from 90° to 0° at intervals of $\Delta\theta = -22.5°$ are shown in Table 1 (for comparison with values of f for $n = 3.5$, values of f for $n = \infty$ are also shown). The data in Table 1 for θ, λ, and f ($n = \infty$) are those used to generate lambda small circles of colatitude in Tamsett (Figure 2) [24] in a SAC employing a far-range approximation to simulate finding the direction to an acoustic source.

Table 1. Simulated values of λ and f (for range nd for which $n = 3.5$, and for comparison also for $n = \infty$) for five azimuthal positions of an acoustic source relative to the direction the head is facing θ for an acoustic source inclined to the horizontal at $\varphi = -30°$.

Circle	θ (Longitude)	λ (Colatitude)	f ($n = 3.5$)	f ($n = \infty$)
1	90.000°	30.000°	0.8636	0.8660
2	67.500°	36.860°	0.7971	0.8001
3	45.000°	52.239°	0.6085	0.6124
4	22.500°	70.645°	0.3284	0.3314
5	00.000°	90.000°	0.0000	0.0000

With the far-range approximation relaxed, a SAC having an additional dimension is demonstrated here to simulate finding range as well as direction to an acoustic source. Having generated simulated measurements of arrival time differences by computing values of f for $n = 3.5$ (using Equation (8); these values populate column 4 in Table 1) we now treat range to the acoustic source as though it were unknown to illustrate how SACs for multiple values of n may be used to optimise the value of n to provide an estimate of range. From the simulated measurements of f for $n = 3.5$, we compute values of λ using Equation (9) for simulated trial values of n ranging from 1.5 to 5.5 (in $\Delta n = 0.5$ intervals). These data populate Table 2 and are used to generate the lambda plots for each of the trial values of n shown in Figure 2.

Table 2. Values of colatitude λ computed from the simulated values of f ($n = 3.5$) in Table 1 and for trial values of range nd for which n varies from 1.5 to 5.5 in intervals of $\Delta n = 0.5$. These data are used to generate the lambda plots as a function of n shown in Figure 2 to illustrate how synthetic aperture calculation can be deployed to estimate range to an acoustic source.

λ	$n = 1.5$	$n = 2.0$	$n = 2.5$	$n = 3.0$	$n = 3.5$	$n = 4.0$	$n = 4.5$	$n = 5.0$	$n = 5.5$
1	28.850°	29.470°	29.754°	29.908°	30.000°	30.060°	30.101°	30.132°	30.152°
2	35.597°	36.276°	36.589°	36.758°	36.860°	36.926°	36.971°	37.002°	37.028°
3	50.994°	51.659°	51.969°	52.137°	52.239°	52.305°	52.350°	52.380°	52.406°
4	69.859°	70.277°	70.473°	70.581°	70.645°	70.688°	70.717°	70.737°	70.753°
5	90.000°	90.000°	90.000°	90.000°	90.000°	90.000°	90.000°	90.000°	90.000°

The lambda plots in Figures 2 and 3 are shown zoomed-in to a small area subtending an azimuth of just two degrees of longitude and an elevation of two degrees of latitude (for comparison, the face of the Moon subtends an angle of ~0.5° at the Earth's surface) so that the differences between the trial values of n can easily be seen. The lambda circles appear as colored lines in Figure 2. A corresponding zoomed-out plot over a wide field of audition is shown in Tamsett (Figure 2) [24].

It is seen in Figure 2 that the lambda plot that generates the best focus at a point is the one for which $n = 3.5$, as we would expect. The lambda plots for $n < 3.5$ form an unfocussed cluster of

points of intersection of lambda circles to the northeast of the known location of the simulated acoustic source, and the lambda plots for $n > 3.5$ form a cluster of points of intersection to the south-west of the location of the source.

It can also be seen in Figure 2 that the plots become more rapidly unfocussed in stepping down in equal steps of n from the optimal value of $n = 3.5$, than in stepping up from the optimal value. This arises because, as in the visual system in which binocular estimates of distance to objects based on the convergence applied to the axes of the eyes to best achieve a focused image are more accurate for shorter range, so too are estimates of range to acoustic sources from SAC in a binaural system more accurate for shorter range.
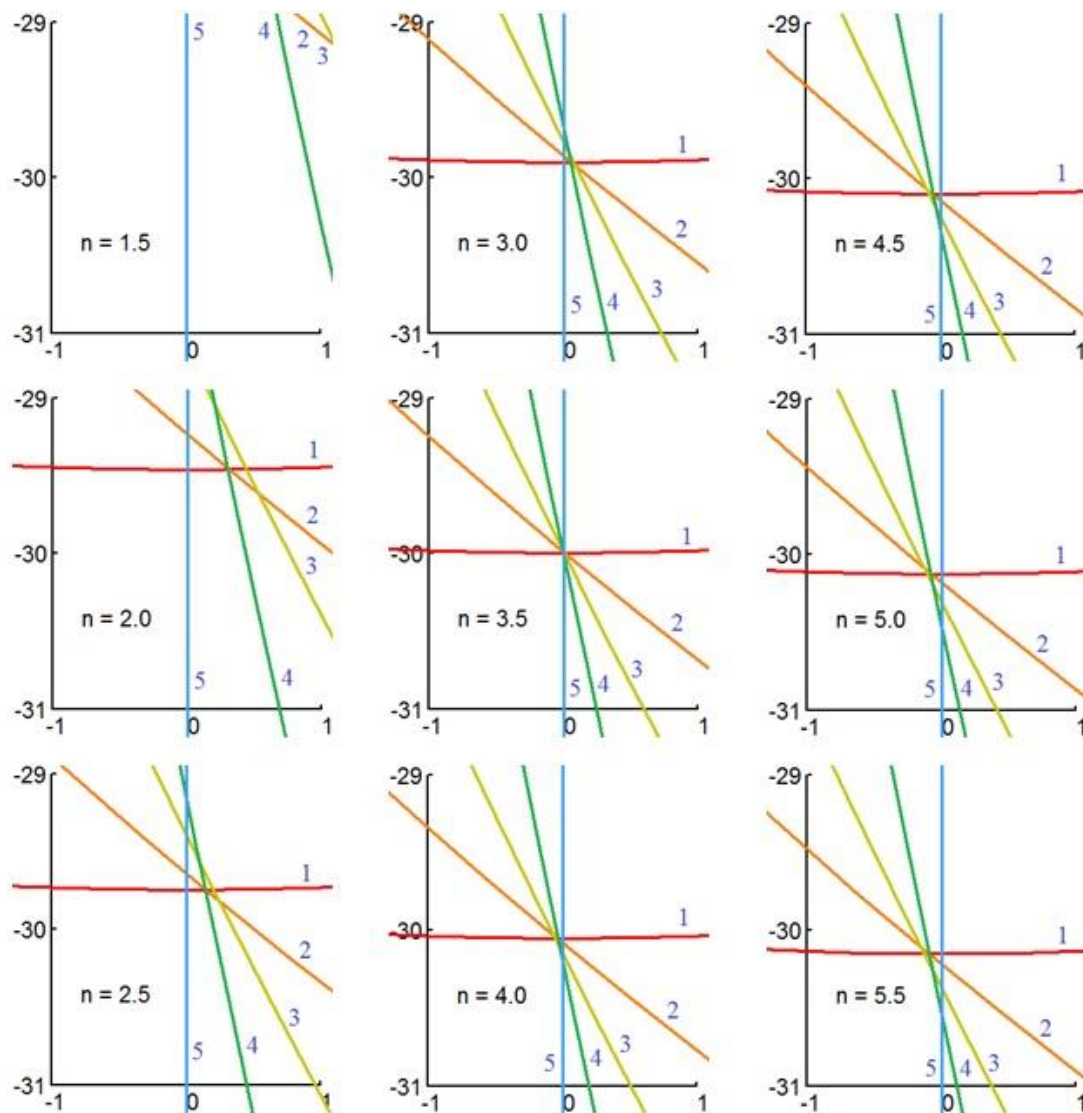
Figure 2. Parts of lambda small circles of colatitude drawn in a virtual field of audition shown as charts in degrees latitude and longitude in Mercator projection for an azimuth of $\theta = 0°$ after the head has turned from $\theta = 90°$ (in $\Delta\theta = -22.5°$ intervals) for an acoustic source at a range of $n = 3.5$ times the distance between the ears, at an elevation of $\varphi = -30°$. The range is treated as unknown, and lambda circles for nine trial values of range nd (where $n = 1.5$ to 5.5) are shown. The set of circles that best intersects to form a focus at a point optimizes n for an estimate of range to an acoustic source.

3.2. Inclined Auditory Axis

An inclination in the auditory axes across the head (an adaptation in the auditory system of species of owl) confers a distinct advantage for unambiguously finding direction to acoustic sources [44]. It is demonstrated in Tamsett [24] that a SAC process acting on data acquired by a binaural system having an inclination in the auditory axes across the head confers an ability to robustly and elegantly disambiguate direction in azimuth and elevation to acoustic sources.

Tables analogous to Tables 1 and 2, for an auditory axis inclined at $i = 20°$ to the right across the head are shown as Tables 3 and 4.

Table 3. Simulated values of λ and f ($n = 3.5$ and ∞) for an auditory system with an auditory axis inclined to the right across the head $i = 20°$, for six azimuthal positions of an acoustic source relative to the direction the head is facing θ for an acoustic source inclined to the horizontal at $\varphi = -30°$.

Circle	θ (Longitude)	λ (Colatitude)	f ($n = 3.5$)	f ($n = \infty$)
1	90.000°	10.000°	0.9845	0.9848
2	67.500°	22.652°	0.9214	0.9229
3	45.000°	41.716°	0.7431	0.7465
4	22.500°	61.155°	0.4787	0.4823
5	00.000°	80.153°	0.1693	0.1710
6	$-12.130°$	90.000°	0.0000	0.0000

Table 4. Values of colatitude λ computed from the simulated values of f ($n = 3.5$) in Table 3 for an auditory system with an auditory axis inclined to the right across the head at $i = 20°$, and for trial values of range nd for which n varies from 1.5 to 5.5 in intervals of $\Delta n = 0.5$. These data are used to generate the lambda plots as a function of n shown in Figure 3.

λ	$n = 1.5$	$n = 2.0$	$n = 2.5$	$n = 3.0$	$n = 3.5$	$n = 4.0$	$n = 4.5$	$n = 5.0$	$n = 5.5$
1	09.537°	09.788°	09.902°	09.963°	10.000°	10.024°	10.040°	10.052°	10.060°
2	21.699°	22.214°	22.449°	22.576°	22.652°	22.701°	22.735°	22.759°	22.777°
3	40.416°	41.115°	41.436°	41.611°	41.716°	41.784°	41.831°	41.864°	41.889°
4	60.080°	60.654°	60.921°	61.067°	61.155°	61.213°	61.252°	61.280°	61.301°
5	79.733°	79.956°	80.061°	80.119°	80.153°	80.176°	80.192°	80.203°	80.211°
6	90.000°	90.000°	90.000°	90.000°	90.000°	90.000°	90.000°	90.000°	90.000°

The angles λ in Table 3 are computed for values of θ, φ, and i [24] from:

$$\lambda = \text{acos}(\sin\theta \cos i \cos\varphi + \sin i \sin\varphi) \tag{11}$$

The angle θ for $\lambda = 90°$, circle 6, in Table 3 for which the inclined median plane (the plane normal to the auditory axis at the auditory center) intersects the source of sound is computed [24] from:

$$\theta = \text{asin}(\tan i \tan\varphi) \tag{12}$$

A figure analogous to Figure 2 is shown in Figure 3. The data in Table 3 for: θ, λ and f ($n = \infty$) are those used to produce the lambda circles shown in a zoomed-out plot over a wide field of audition in Tamsett (Figure 4) [24].

Whilst an inclination of the auditory axis confers an advantage over a horizontal auditory axis for direction finding, no advantage is gained for the purpose of range finding.
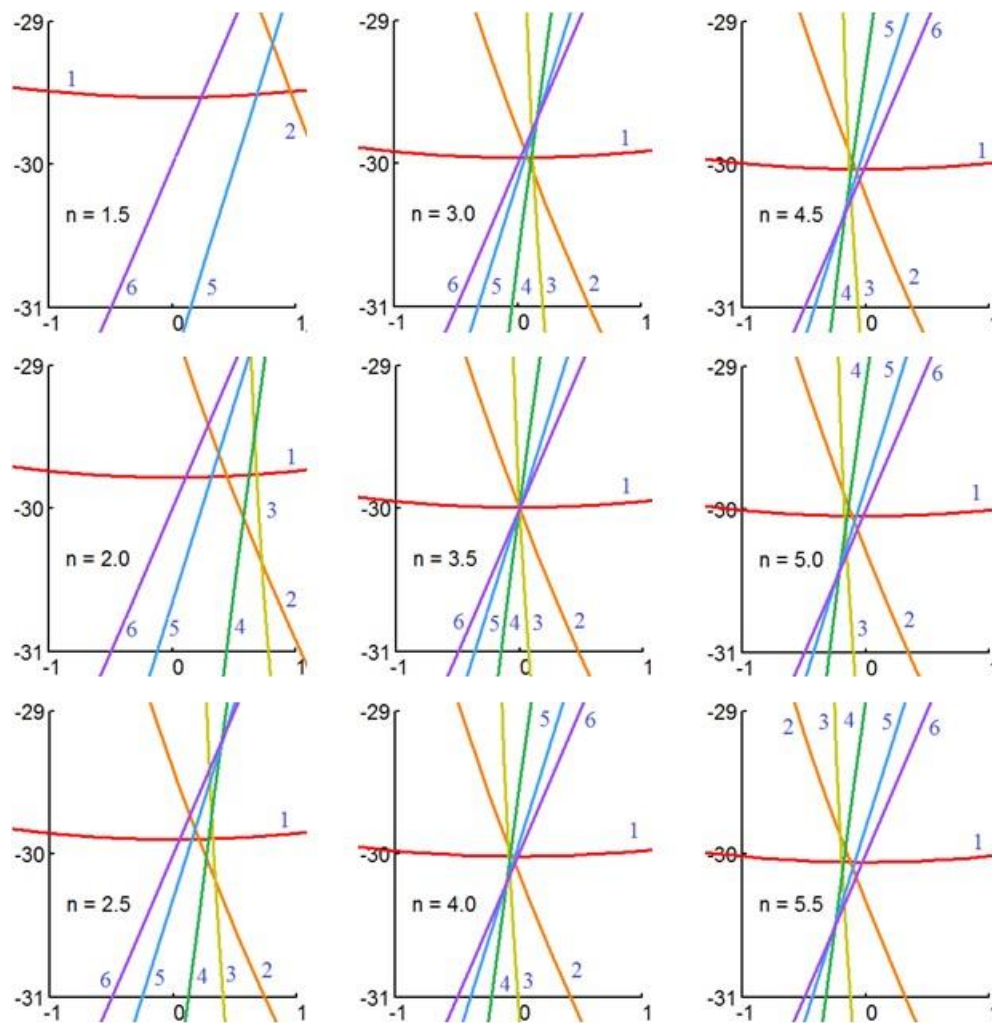
Figure 3. Parts of acoustic lambda small circles of colatitude in a virtual field of audition shown as charts in degrees latitude and longitude in Mercator projection for an auditory axis inclined at $i = 20°$ to the right across the head, for an azimuth of $\theta = -12.1°$ after the head has turned from $\theta = 90°$ (in $\Delta\theta = -22.5°$ intervals) for an acoustic source at a range of $n = 3.5$ times the distance between the ears, at an elevation of $\varphi = -30°$. The range is treated as unknown and lambda circles for nine trial values of range nd are shown. The set of circles that best intersects to form a focus at a point optimizes n for an estimate of range to acoustic source.

4. Discussion

The method described for estimating range to an acoustic source could in principle form the basis of an implementation in a binaural robotic system, however the challenge in an implementation is the need for very accurate measurement of arrival time/level differences between the antennae to achieve the spatial resolution required to successfully discriminate the effect of range in a SAC for an estimate of range to be made.

It is apparent from Figures 2 and 3 that high spatial resolution of acoustic location better than $1/4°$ would appear to be necessary, to achieve an estimate of a range even as short as 3.5-times the distance between the ears, to an accuracy no better than ~30%. This is beyond human capability which can locate direction with a resolution estimated to be $1.0°$ [48] to $1.5°$ [26] in the direction the head is facing, and furthermore would also appear to be beyond the capability of owls estimated to be able to resolve direction to approximately only $3°$ [49,50] despite being able to catch prey in total darkness from audition alone [44].

However, it might be possible in principle for the auditory capability of a robotic system to exceed those of humans and owls for estimates of range based on the three-dimensional SAC method described and demonstrated here. Highly accurate estimates of difference in arrival times at the ears, better than 0.5% of the travel time for the distance between the ears, are required and an obvious ploy to adopt to attempt to achieve this in a robotic system intended to explore the possibilities of the method would be to allow the distance between the listening antennae to be large, at least in the first instance. Because differences in arrival times are related to differences in acoustic ray path lengths via the acoustic transmission velocity, the velocity of sound will also be required to be known to a similar degree of accuracy.

It will be possible to calibrate an auditory system's estimate of the acoustic transmission velocity by performing a three-dimensional (3D) SAC analogous to the one described here for estimating range, but instead listening to a sound from a source at a sufficiently large range to allow the far-range approximation to be made, and optimizing the acoustic transmission velocity for the strongest response in a 3D SAC.

Any implementation in nature of acoustic range estimation based on the SAC principle will incorporate biological neural-ware components to achieve the equivalent of mathematical operations in the software/firmware components of a robotic system.

5. Summary

A method for determining the direction to an acoustic source with a pair of omnidirectional listening antennae in a robotic system, or by an animal's ears, as the head is turned is described in Tamsett [24] in which a measure of acoustic energy received at the antennae/ears is extended over lambda circles of colatitude and integrated in a virtual/subconscious acoustic image of the field of audition. This constitutes a synthetic aperture computation (SAC) analogous to data processes in anthropic synthetic aperture radar, sonar, and seismic [42] technologies. The method has been extended in this paper and a far range approximation [24] relaxed to allow SAC as a function of range. By optimizing range to find maxima in integrated energy over multiple sets of lambda circles generated as a function of range; or alternatively, a best focus to a point of intersection of lambda circles, range as well as direction to acoustic sources can in principle be estimated.

This embellished SAC process promotes the direction finding capability of a pair of antennae to that of a large two-dimensional stationary array of acoustic antennae, with not only beam forming direction finding but also range finding capability. The method appears to be beyond the acoustic direction finding capabilities of human audition, however it might nevertheless find utility in a binaural robotic system capable of sufficiently accurate measurement of arrival time/level differences between the listening antennae.

Conflicts of Interest: The author declares there to be no conflict of interest.

References

1. Lollmann, H.W.; Barfus, H.; Deleforge, A.; Meier, S.; Kellermann, W. Challenges in acoustic signal enhancement for human-robot communication. In Proceedings of the ITG Conference on Speech Communication, Erlangen, Germany, 24–26 September 2014.
2. Takanishi, A.; Masukawa, S.; Mori, Y.; Ogawa, T. Development of an anthropomorphic auditory robot that localizes a sound direction. *Bull. Cent. Inform.* **1995**, *20*, 24–32. (In Japanese)
3. Matsusaka, Y.; Tojo, T.; Kuota, S.; Furukawa, K.; Tamiya, D.; Nakano, Y.; Kobayashi, T. Multi-person conversation via multi-modal interface—A robot who communicates with multi-user. In Proceedings of the 16th National Conference on Artificial Intelligence (AAA1-99), Orlando, FL, USA, 18–22 July 1999; pp. 768–775.

4.　Ma, N.; Brown, G.J.; May, T. Robust localisation of multiple speakers exploiting deep neural networks and head movements. In Proceedings of the INTERSPEECH 2015, Dresden, Germany, 6–10 September 2015; pp. 3302–3306.

5.　Schymura, C.; Winter, F.; Kolossa, D.; Spors, S. Binaural sound source localization and tracking using a dynamic spherical head model. In Proceedings of the INTERSPEECH 2015, Dresden, Germany, 6–10 September 2015; pp. 165–169.

6.　Winter, F.; Schultz, S.; Spors, S. Localisation properties of data-based binaural synthesis including translator head-movements. In Proceedings of the Forum Acusticum, Krakow, Poland, 31 January 2014.

7.　Bustamante, G.; Portello, A.; Danes, P. A three-stage framework to active source localization from a binaural head. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane, Australia, 19–24 April 2015; pp. 5620–5624.

8.　May, T.; Ma, N.; Brown, G. Robust localisation of multiple speakers exploiting head movements and multi-conditional training of binaural cues. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane, Australia, 19–24 April 2015; pp. 2679–2683.

9.　Ma, N.; May, T.; Wierstorf, H.; Brown, G. A machine-hearing system exploiting head movements for binaural sound localisation in reverberant conditions. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brisbane, Australia, 19–24 April 2015; pp. 2699–2703.

10.　Bhadkamkar, N.A. Binaural source localizer chip using subthreshold analog CMOS. In Proceedings of the IEEE International Conference on Neural Networks, Orlando, FL, USA, 28 June–2 July 1994; Volume 3, pp. 1866–1870.

11.　Willert, V.; Eggert, J.; Adamy, J.; Stahl, R.; Korner, E. A probabilistic model for binaural sound localization. *IEEE Trans. Syst. Man Cybern. B* **2006**, *36*, 982–994. [CrossRef]

12.　Voutsas, K.; Adamy, J. A biologically inspired spiking neural network for sound source lateralization. *IEEE Trans. Neural Netw.* **2007**, *18*, 1785–1799. [CrossRef] [PubMed]

13.　Liu, J.; Perez-Gonzalez, D.; Rees, A.; Erwin, H.; Wermter, S. A biologically inspired spiking neural network model of the auditory midbrain for sound source localization. *Neurocomputing* **2010**, *74*, 129–139. [CrossRef]

14.　Nakadai, K.; Lourens, T.; Okuno, H.G.; Kitano, H. Active audition for humanoids. In Proceedings of the 17th National Conference Artificial Intelligence (AAAI-2000), Austin, TX, USA, 30 July–3 August 2010; pp. 832–839.

15.　Cech, J.; Mittal, R.; Delefoge, A.; Sanchez-Riera, J.; Alameda-Pineda, X. Active speaker detection and localization with microphone and cameras embedded into a robotic head. In Proceedings of the IEEE-RAS International Conference on Humanoid Robots (Humanoids), Atlanta, GA, USA, 15–17 October 2013; pp. 203–210.

16.　Deleforge, A.; Drouard, V.; Girin, L.; Horaud, R. Mapping sounds on images using binaural spectrograms. In Proceedings of the European Signal Processing Conference, Lisbon, Portugal, 1–5 September 2014; pp. 2470–2474.

17.　Nakamura, K.; Nakadai, K.; Asano, F.; Ince, G. Intelligent sound source localization and its application to multimodal human tracking. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, San Francisco, CA, USA, 25–30 September 2011; pp. 143–148.

18.　Yost, W.A.; Zhong, X.; Najam, A. Judging sound rotation when listeners and sounds rotate: Sound source localization is a multisystem process. *J. Acoust. Soc. Am.* **2015**, *138*, 3293–3310. [CrossRef] [PubMed]

19.　Kim, U.H.; Nakadai, K.; Okuno, H.G. Improved sound source localization in horizontal plane for binaural robot audition. *Appl. Intell.* **2015**, *42*, 63–74. [CrossRef]

20.　Rodemann, T.; Heckmann, M.; Joublin, F.; Goerick, C.; Scholling, B. Real-time sound localization with a binaural head-system using a biologically-inspired cue-triple mapping. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Beijing, China, 9–15 October 2006; pp. 860–865.

21.　Portello, A.; Danes, P.; Argentieri, S. Acoustic models and Kalman filtering strategies for active binaural sound localization. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, San Francisco, CA, USA, 25–30 September 2011; pp. 137–142.

22.　Sun, L.; Zhong, X.; Yost, W. Dynamic binaural sound source localization with interaural time difference cues: Artificial listeners. *J. Acoust. Soc. Am.* **2015**, *137*, 2226. [CrossRef]

23.　Zhong, X.; Sun, L.; Yost, W. Active binaural localization of multiple sound sources. *Robot. Autom. Syst.* **2016**, *85*, 83–92. [CrossRef]

24.　Tamsett, D. Synthetic aperture computation as the head is turned in binaural direction finding. *Robotics* **2017**, *6*. [CrossRef]

25.　Wightman, F.L.; Kistler, D.J. The dominant role of low frequency interaural time differences in sound localization. *J. Acoust. Soc. Am.* **1992**, *91*, 1648–1661. [CrossRef] [PubMed]

26. Brughera, A.; Danai, L.; Hartmann, W.M. Human interaural time difference thresholds for sine tones: The high-frequency limit. *J. Acoust. Soc. Am.* **2013**, *133*. [CrossRef] [PubMed]

27. Moore, B. *An Introduction to the Psychology of Hearing*; Academic Press: San Diego, CA, USA, 2003.

28. Sayers, B.M.; Cherry, E.C. Mechanism of binaural fusion in the hearing of speech. *J. Acoust. Soc. Am.* **1957**, *36*, 923–926. [CrossRef]

29. Jeffress, L.A.A. A place theory of sound localization. *J. Comp. Physiol. Psychol.* **1948**, *41*, 35–39. [CrossRef] [PubMed]

30. Colburn, H.S. Theory of binaural interaction based on auditory-nerve data. 1. General strategy and preliminary results in interaural discrimination. *J. Acoust. Soc. Am.* **1973**, *54*, 1458–1470. [CrossRef] [PubMed]

31. Kock, W.E. Binaural localization and masking. *J. Acoust. Soc. Am.* **1950**, *22*, 801–804. [CrossRef]

32. Durlach, N.I. Equalization and cancellation theory of binaural masking-level differences. *J. Acoust. Soc. Am.* **1963**, *35*, 1206–1218. [CrossRef]

33. Licklider, J.C.R. Three auditory theories. In *Psychology: A Study of a Science*; Koch, S., Ed.; McGraw-Hill: New York, NY, USA, 1959; pp. 41–144.

34. Smith, P.; Joris, P.; Yin, T. Projections of physiologically characterized spherical bushy cell axons from the cochlear nucleus of the cat: Evidence for delay lines to the medial superior olive. *J. Comp. Neurol.* **1993**, *331*, 245–260. [CrossRef] [PubMed]

35. Brand, A.; Behrend, O.; Marquardt, T.; McAlpine, D.; Grothe, B. Precise inhibition is essential for microsecond interaural time difference coding. *Nature* **2002**, *417*, 543–547. [CrossRef] [PubMed]

36. Roffler, S.K.; Butler, R.A. Factors that influence the localization of sound in the vertical plane. *J. Acoust. Soc. Am.* **1968**, *43*, 1255–1259. [CrossRef] [PubMed]

37. Batteau, D. The role of the pinna in human localization. *Proc. R. Soc. Lond. Ser. B* **1967**, *168*, 158–180. [CrossRef]

38. Middlebrooks, J.C.; Makous, J.C.; Green, D.M. Directional sensitivity of sound-pressure levels in the human ear canal. *J. Acoust. Soc. Am.* **1989**, *86*, 89–108. [CrossRef] [PubMed]

39. Rodemann, T.; Ince, G.; Joublin, F.; Goerick, C. Using binaural and spectral cues for azimuth and elevation localization. In Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, Nice, France, 22–26 September 2008; pp. 2185–2190.

40. Wallach, H. The role of head movement and vestibular and visual cues in sound localisation. *J. Exp. Psychol.* **1940**, *27*, 339–368. [CrossRef]

41. Perrett, S.; Noble, W. The effect of head rotations on vertical plane sound localization. *J. Acoust. Soc. Am.* **1997**, *102*, 2325–2332. [CrossRef] [PubMed]

42. Lurton, X. Seafloor-mapping sonar systems and Sub-bottom investigations. In *An Introduction to Underwater Acoustics: Principles and Applications*, 2nd ed.; Springer: Berlin, Germany, 2010; pp. 75–114.

43. Rice, J.J.; May, B.J.; Spirou, G.A.; Young, E.D. Pinna-based spectral cues for sound localization in cat. *Hear. Res.* **1992**, *58*, 132–152. [CrossRef]

44. Payne, R.S. Acoustic location of prey by barn owls. *J. Exp. Biol.* **1971**, *54*, 535–573. [PubMed]

45. Coleman, P.D. Failure to Localize the Source Distance of an Unfamiliar Sound. *J. Acoust. Soc. Am.* **1962**, *34*, 345–346. [CrossRef]

46. Plenge, G. On the problem of "in head localization". *Acustica* **1972**, *26*, 213–221.

47. Stern, R.; Brown, G.J.; Wang, D.L. Binaural sound localization. In *Computational Auditory Scene Analysis*; Wang, D.L., Brown, G.L., Eds.; John Wiley and Sons: Chichester, UK, 2005; pp. 1–34.

48. Mills, A.W. On the minimum audible angle. *J. Acoust. Soc. Am.* **1958**, *30*, 237–246. [CrossRef]

49. Bala, A.D.S.; Spitzer, M.W.; Takahashi, T.T. Prediction of auditory spatial acuity from neural images of the owl's auditory space map. *Nature* **2003**, *424*, 771–774. [CrossRef] [PubMed]

50. Knudsen, E.I.; Konishi, M. Mechanisms of sound localization in the barn owl (*Tyto alba*). *J. Comp. Physiol. A* **1979**, *133*, 13–21. [CrossRef]