Acta Acustica 2021, **5**, 45

https://doi.org/10.1051/aacus/2021039

Available online at:
https://acta-acustica.edpsciences.org

Topical Issue - Auditory models: from binaural processing to multimodal cognition

REVIEW ARTICLE OPEN ACCESS

Towards modelling active sound localisation based on Bayesian inference in a static environment

Glen McLachlan[1,*], Piotr Majdak[2], Jonas Reijniers[1], and Herbert Peremans[1]

[1] Department of Engineering Management, University of Antwerp, 2000 Antwerp, Belgium
[2] Acoustics Research Institute, Austrian Academy of Sciences, 1040 Vienna, Austria

Received 19 April 2021, Accepted 20 September 2021

Abstract – Over the decades, Bayesian statistical inference has become a staple technique for modelling human multisensory perception. Many studies have successfully shown how sensory and prior information can be combined to optimally interpret our environment. Because of the multiple sound localisation cues available in the binaural signal, sound localisation models based on Bayesian inference are a promising way of explaining behavioural human data. An interesting aspect is the consideration of dynamic localisation cues obtained through self-motion. Here we provide a review of the recent developments in modelling dynamic sound localisation with a particular focus on Bayesian inference. Further, we describe a theoretical Bayesian framework capable to model dynamic and active listening situations in humans in a static auditory environment. In order to demonstrate its potential in future implementations, we provide results from two examples of simplified versions of that framework.

Keywords: Sound localisation, Active listening, Dynamic cues, Bayes, Models

1 Introduction

Sound localisation is a primary function of the human auditory system. Besides the well established evolutionary advantages [1], it is a crucial process for attention control and self-orientation. Proper understanding and implementation of the cues responsible for localisation is relevant for a range of modern audio applications, such as binaural hearing aids and three-dimensional audio displays for augmented or virtual reality [2].

The binaural nature of the auditory system is of high importance for localisation of the lateral position (Fig. 1) of a sound source [3]. Humans obtain information about the source through the interaural differences in time of arrival (interaural time difference, ITD) and level (interaural level difference, ILD). However, the ITD and ILD cues do not provide enough information for accurate localisation beyond the horizontal plane, as several source locations will give rise to nearly the exact same binaural cues in the so called "cones of confusion" [4]. Monaural spectral cues, which result from the filtering properties of the outer ear, head and torso, carry additional information on the polar position of the source (Fig. 1). This spectral information aids in resolving the ambiguity in the binaural cues [5].

The aforementioned ITD, ILD and spectral cues can be considered "static", as they are usually obtained in a situation where neither the source, nor the head undergoes any movement. However, in addition to these static cues, the auditory system can also utilise "dynamic" cues, which are obtained by either sound source or head movement. Thus, dynamic cues can be defined as the changes in static cues during motion. Dynamic cues are beneficial during sound localisation, especially for resolving front–back confusions (e.g., [6]). They aid localisation in some way, but their importance relative to static cues is still an active point of research. Furthermore, dynamic cues obtained from self motion, e.g., head movements, bring the additional challenge of processing sensorimotor information. As a result, the majority of state of the art models for sound localisation do not include the use of head movements [7].

There are many available models, which each focus on a specific aspect of binaural localisation, such as processing of binaural cues [8, 9], spectral cues [10, 11], or reverberant environments [12]. Over the past decade, machine learning techniques have also been applied to the modelling problem (e.g., [13, 14]). Despite promising results, these techniques require substantial amounts of training data and can be difficult to understand due to their black-box nature [15].

Bayesian inference is a method to optimally combine information about a multivariate system, when relying on noisy observations only. Bayesian inference has often been shown to take place not only in human multisensory perception [16–20] but also human perception based on multiple cues within a single modality [21–23]. Because of the

*Corresponding author: glen.mclachlan@uantwerpen.be

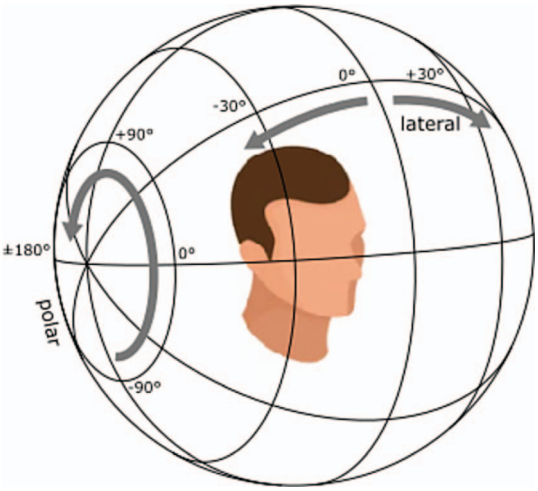

Figure 1. Interaural-polar coordinate system as used in [10, 26, 27]. The lateral angle of 0° describes sources located on the median plane. The polar angle of 0° describes sources located on the horizontal plane at the eye level. Note that in this system, the lateral angle increases to the left, providing the advantage that for sources located at the eye level, the lateral angle coincides with the azimuth angle of the widely used spherical coordinate system.

multiple sound localisation cues available in the binaural signal, sound localisation models based on Bayesian inference seem to be a promising way of explaining behavioural human data. Temporal integration and learning can be modelled through recursive Bayesian estimation, where probabilities and estimates are updated recursively over time with incoming measurements [24, 25].

This article has two main purposes. First, we review the relevance of dynamic cues and their role in existing models of sound localisation, with a particular focus on the implementation of Bayesian inference. Second, we describe a recursive theoretical framework for dynamic listening through Bayesian inference. This framework aims at modelling dynamic listening situations which involve stationary sound sources in combination with head movements.

2 Static listening

2.1 Acoustic features and perceptual cues

Sound source localisation consists of determining the position of a source in three dimensions comprising two angles and the distance. In the interaural-polar coordinate system, the two angles are defined as the lateral and polar angles relative to a single pole passing through the two ears, i.e., the interaural axis. The interaural-polar system as used in [10, 26, 27] with a fixed distance between the listener and the source is illustrated in Figure 1. The static physical acoustic cues for localisation are captured by the (binaural) head-related transfer functions (HRTFs), which describes the filtering of the sound for a given direction by the listener's anatomy as recorded at the two ear drums.

For sound-source localisation along the lateral angle, the two main cues are the ITDs and the ILDs which are caused by the wave propagation time difference and the shadowing effects of the head, respectively [3]. ITD as a function of lateral angle roughly follows a sine shape, with zeroes on the median plane and maxima on the interaural axis [28]. This means that small displacements around the median plane produce larger changes in ITD than the same displacements at the lateral sides of the head. The ILD calculated for a spherical head model doesn't show maxima on the interaural axis, but for locations 45° on either side of that axis [28]. For narrowband sounds, the ILD cues are dominant at the middle to high-frequency range of human hearing, and the ITD cues are particularly important for low frequencies [29]. This is known as the duplex theory of sound localisation. For broadband sounds, which encompasses most natural signals, both cues have substantial weight, but ITD dominates for most listeners [30].

ITDs and ILDs produced by a sound at one location are ambiguous cues as they can also be produced by a sound at any location on the surface of a cone centred on the interaural axis, an effect known as the "cone of confusion" [3]. Thus, in addition to these interaural broadband cues, the asymmetric and convoluted shape of the outer ears functions as a direction-dependent filter by causing frequency-dependent interference before sound waves reach the ear drums. The spectral cues introduced at each ear provide spatial information along the sagittal planes that helps to disambiguate the cones of confusion [5] which results in smaller elevation errors and a reduced so-called quadrant error rate, i.e., rate of confusing the spatial quadrant of the source direction, including the confusions between front and back and top and bottom [26, 31]. Thus, the interaural-polar coordinate system provides a simple but complete representation of all sound directions from the perceptual perspective [32], in which the lateral angle depends mostly on interaural cues and polar angle depends mostly on monaural spectral cues (see Fig. 1).

Despite the varying contributions from different spectral regions, incoming sound must comprise sufficient energy in the relevant frequency region to make use of the spectral cues by the auditory system [33, 34]. The human pinna's most prominent spectral notch related to the sound's polar angle falls within the 6–9 kHz band, which varies systematically and monotonically with polar angle [34]. Acoustic features above 9 kHz still dependent on the polar angle, but they vary in a much more complex way. As a general upper limit, frequencies up to l6 kHz are evaluated by the auditory system in order to localise the direction of a sound [35]. As for the lower frequency limit, sounds below 4 kHz have wavelengths that are too large to be affected by the dimensions of the pinnae and, thus, the resonances are direction independent [36]. Additionally, the effectiveness of monaural cues is highly listener specific, due to individual head and ear morphology [37]. This is a prominent issue in 3D auditory displays for sound presentation over headphones [38] as such systems require listener-specific HRTFs to reproduce the spectral cues with full accuracy.

Note that in this article we focus on the direction, and put less attention on the third dimension, the distance. On the one hand, distance perception is closely linked to sound reverberation [39]. On the other hand, our proposed framework is expandable to consider additional variables and more complex problems. For example, in the near field (distances below 1 m), ILDs become a significant cue for the disambiguation of source location [40] and this information could be used to extend our considerations to more complex localisation scenarios. Also, auditory motion parallax can be exploited to assess the relative distances of two sound sources [41].

2.2 Ill-posed problem and prior information

Even for sound sources that meet the requirements above, polar angle estimation is still argued to be a mathematically ill-posed problem [42], as the spectrum of the signal at the eardrum results from a time-domain convolution of two unknowns: the actual source spectrum and the particular direction-dependent HRTF. This means that a priori knowledge of the source spectrum and/or direction helps to differentiate between spectral cues resulting from the source properties and from the filtering by the pinnae. Thus, a listener with an a priori knowledge is better able to estimate the pinna filtering characteristics from an incoming sound and associate those characteristics with the appropriate source position.

Despite the ill-posed problem for sound's polar-angle estimation, human localisation performance can be accurate and precise for most sound directions. Thus, to estimate the most likely direction, the auditory system seems to complement the acoustic cues with non-acoustic information about sounds and the environment. For example, the auditory system considers certain parts of the sensory information to be more reliable than others, such as different weightings on different frequency bands [34]. With respect to sound localisation, priors emphasising the central directions helped to describe the systematic underestimation of peripheral source directions in owls [43]. There also appears to be a clear mapping between frequency and elevation estimation, where high pitch is consistently mapped to high positions and vice versa [44]. A priori assumptions such as the HRTF being unique for each sound elevation and natural source spectra not resembling HRTFs helped in modelling the process of sound localisation [45].

Interestingly, sound-localisation mechanisms seem to be independent along the horizontal and vertical dimensions, providing evidence that they may be embedded as distinct strategies to deal with spatial uncertainty in the acoustic environment [42]. In the same study, azimuth estimation did not require a prior. Conversely, elevation estimation did require a prior in the form of a Gaussian spatial distribution centred around the horizontal plane. This is in line with the current understanding of multisensory perception where priors are independently encoded [46]. The elegant inclusion of such prior information is a major advantage of a Bayesian framework.

In vision, priors have been discovered, e.g., observers tend to underestimate the speed of an object as they initially assume them to move stationary or move slowly [47]. These priors may also apply to audition. In fact, an analogous prior for low velocity of auditory sources has already been suggested [48, 49].

3 Dynamic listening

Our acoustic environment is in constant motion, due to both animate objects and listener movements. This makes the dynamic listening problem twofold [50]: (1) How is motion perceived and encoded in the auditory system? (2) How can a listener disambiguate moving sources from the apparent motion caused by head rotation? Both questions will be addressed in this section.

It is important to distinguish here between "passive" dynamic listening and "active" dynamic listening, particularly because several definitions exist in different fields related to audition [51–54]. According to our definition, passive dynamic listening involves a dynamic acoustic environment without the employment of head movements, i.e., dynamic cues are solely produced by moving sound sources. In contrast, in the active dynamic listening situation listeners can rotate their heads and obtain dynamic cues even from static sound sources.

There are three degrees of freedom considered in most research related to dynamic listening: head rotation around the z-axis (yaw), head turn around the y-axis (pitch) and head tilt around the x-axis (roll), see Figure 2. Note that in this article, we focus on the directional localisation process, thus we do not consider head translations, which are usually related to distance estimation [55].

3.1 Acoustic features and perceptual cues

Wallach first suggested that dynamic ITDs and ILDs associated with head rotation are used to refine localisation accuracy, especially along the cones of confusion [56]. He argued that head yaw rotations would eliminate front–back ambiguity due to the contrasting change in the interaural cues provided by a stationary source [57]. The contribution of dynamic cues to resolve front from back has since been empirically shown multiple times [58–60], especially in conditions in which spectral cues are not fully accessible to the auditory system [38, 61, 62]. Dynamic interaural cues contribute more to front–back resolution than dynamic monaural spectral cues [63] and dynamic ITD is a more salient cue than dynamic ILD [57].

Not only yaw rotations produce a strong dynamic cue, head rolls provide supplementary information to resolve up-down confusions [36]. The contribution of yaw and roll to the process of sound localisation based on ITD can be investigated with the so-called ITD angular rate, i.e., $\mathrm{dITD}/\mathrm{d}\alpha$ with α being the source angle along a given rotation axis. ITD angular rate describes the change in ITD caused by the change in the source direction [64]. Figure 3

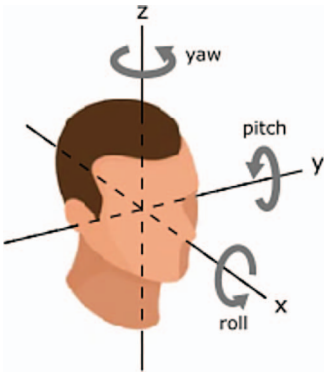

Figure 2. Three degrees of freedom in head orientation: yaw (head rotation), roll (head pivot), and pitch (head tip).

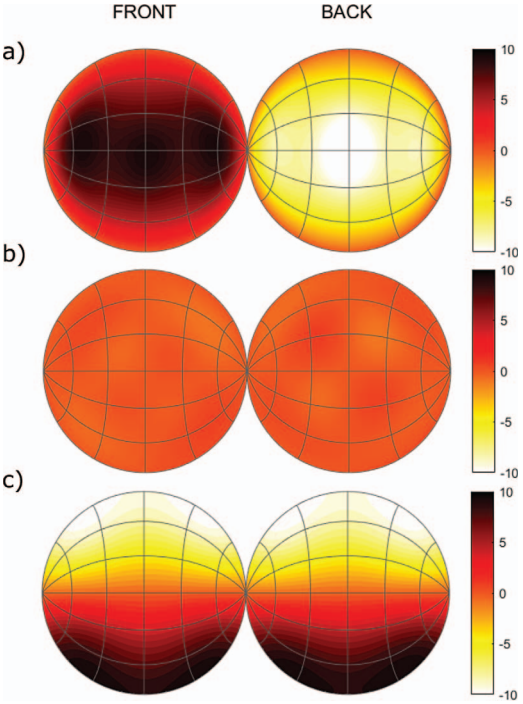

Figure 3. ITD angular rates (dITD/dα) in µs per degree. Here dα corresponds with a positive rotation around the (a) *z*-axis (yaw), (b) *y*-axis (pitch), and (c) *x*-axis (roll). Left and right panels represent source locations in the front and back of the head, respectively.

shows ITD angular rates for the three rotation axes and sources placed over the full sphere. The ITDs were calculated from the HRTFs of a mannequin (KU 100, Neumann, Germany) available from the THK SOFA database [65]. Figure 3 shows that yaw and roll induce large ITD rates providing a strong cue to resolve the cone of confusion. The head pitch, on the other hand, does not seem to evoke significant ITD rates, i.e., it does not provide dynamic cues to sufficiently resolve the cone of confusion.

Dynamic cues also help in estimating the elevation of a sound source. As it can be deduced from Figure 3a, the ITD

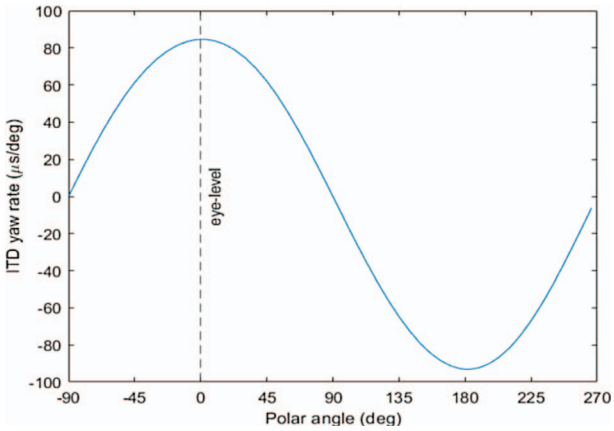

Figure 4. ITD yaw rates (in µs per degree) calculated for the median plane as a function of the sound-source polar angle. Note the sinusoidal shape indicating that the ITD yaw rates do not change linearly with polar angle, showing the largest values at eye level (polar angle of zero), but the largest changes above or below the listener (polar angle of ±90).

angular rates caused by head yaw depend on the source elevation angle [56]. They are large for sources placed on the horizontal plane and nearly zero for those placed directly above or below a listener. This relation is shown in more detail in Figure 4, which shows the ITD angular rate for head yaw (dITD/dα with α being the lateral angle) as a function of the polar angle for sources located on the median plane. The auditory system is able to evaluate these ITD rate differences and associate them with the sound elevation [6, 7, 38, 56, 61]. The sensitivity of this feature varies with the elevation. In reference to Figure 4, the ITD yaw rate is largest for sources on the horizontal plane and the smallest for sources above and below the listener. The opposite is true for the slope of the ITD yaw rate, i.e., the slope is steeper for higher polar angles. The steepness of the slope may explain why elevation estimation based on dynamic ITD only improves for elevations greater than 30° above or below the horizontal plane [6]. Generally, the relation between the dynamic ITD rate and elevation perception seems to be quite complex, as it seems to further depend on the stimulus bandwidth [61] and might even be supported by dynamic spectral cues [38], but not in a monaural listening situation [66].

Head movements do not always improve localisation. Brief sounds played during an ongoing head movement may even degrade localisation accuracy [67, 68]. During rapid head turns, the auditory space can be perceived as distorted or "smeared" [69], which indicates that rotation speed may be a relevant parameter in the process of sound localisation. Interestingly, those distortions only occurred when the sounds were presented near the end of the head turn indicating a complex interaction between the head rotation and perceived auditory space. On top of that, all the spatial cues can vary temporally and our brains need to integrate the information somehow in order to obtain a stable image of the environment. Unfortunately, it is not completely clear yet how the brain accomplishes this task [70].

So far, there is no evidence for peripheral neurons sensitive to auditory motion [71–73], in contrast to those found in the visual system. Still, humans are able to faithfully track a sound's unpredictable movements in the horizontal plane with smooth-pursuit responses of the head, which in turn supports the existence of a pursuit system for auditory head-tracking [74]. This is supported by neurons in the midbrain (inferior colliculus and medial geniculate nucleus) sensitive to dynamic cue changes. This suggests the existence of a higher-level neural network estimating sound motion, similar to that of third-order (acceleration) motion detectors found in vision for cats [75], bats [76], guinae pigs [77], and barn owls [78, 79]. These networks are heavily modulated by attention [80] and have been measured by means of electroencephalography (EEG) [81], providing further evidence for higher cognitive processes involved in decoding sound velocity in humans. Taken together, sound motion is most probably tracked by sampling the estimated source position and integrating that information by higher stages of the auditory system [50, 82], rather than by a continuous measurement of sound velocity in the peripheral stages.

It is generally accepted that the auditory system depends on a type of "temporal integration" [83]. It is important here to distinguish between the operation of mathematical integration (as the term "temporal integration" seems to imply) and the actual process in the analysis of time-variant information. In cognitive sciences, temporal integration considers a variety of models working on various time scales [84]. For example, in the "multiple looks" model, samples or "looks" are taken from the acoustic features, stored in memory, and can be selectively accessed and processed [85]. When applied to the process of sound localisation, the auditory system seems to integrate acoustic information over a duration of approximately 5 ms to form a single look, which are then combined through a leaky integrator, with a stable composite estimation requiring a stimulus duration of approximately 80 ms [86].

More specifically, static elevation estimation seems to require 40–80 ms of broadband input [86]. For static lateralisation, stable performance can be achieved with stimuli as short as 3 ms [87]. During dynamic sound-source localisation, the sound localisation system seems to require a minimum 100 ms of input to yield an improved estimate (likely due to the process of vestibular-auditory integration), with the stimulus duration above 100 ms further improving the localisation performance [6, 57].

Doppler shift, i.e., the frequency shift caused by the motion of the sound source and/or the listener, is an additional dynamic cue that must be noted. Interestingly, within a single frequency band, the binaural Doppler equation results in mathematically equivalent results as the ITD angular rate [88]. Despite its implementation in robotic systems [89], there seems to be no evidence that human use the Doppler effect to localise sound sources. When considering moving sources, however, humans are indeed able to utilise the Doppler shift as a cue for velocity discrimination [90].

3.2 Integration of sensorimotor information

Listeners are capable of dissociating self-motion from source motion, with the largest apparent difference being the additional sensory feedback from the vestibular and proprioceptive systems in the case of self motion [50]. Consequently, the contribution of self-motion implies the consideration of sensorimotor information in modelling the localisation process.

In the human auditory system, acoustic cues are encoded in an egocentric representation, i.e., head-centred reference frame [87]. In the process of spatial inference, the frame of reference needs to be transformed from egocentric to allocentric, i.e., world-centred information about the environment [91]. The auditory system is able to compensate for head rotations during the perception of sound-source motion, though this compensation seems to be incomplete [49]. Complementary information from other senses, integrated with the acoustic input can help to better estimate the allocentric spatial properties of the environment. In fact, mechanisms responsible for building an allocentric frame of reference are based on multisensory processing [92–94].

Thus, it is not surprising that in the process of sound localisation, information needs to be integrated from many systems such as the vestibular system, proprioception or from efference copies of motor commands [95]. For example, performance in a dynamic spatial auditory task improved when dynamic cues were generated by self-induced head motion rather than by the source itself [96]. However, front–back confusions that initially did not resolve with source movement were in fact resolved when source movements were controlled by the listener [58], suggesting that head movements may not be required to produce dynamic cues to resolve front–back ambiguity and, instead, that the listener's priors, e.g., additional information on the direction of the source, contribute strongly.

The extent to which the various information channels contribute to the process of spatial calibration remains an open question. There is a strong indication that vestibular-auditory integration takes place in the sound localisation process as indicated by the requirement of a long stimulus duration (in the range of 100 ms) for an effective use of dynamic auditory cues [57]. However, in that study, listeners rotated their head at a constant velocity and the stimulus was played when the head orientation entered a selected spatial window. Because of this, movement initiation and acceleration, in which sensorimotor mechanisms may play a prominent role, were not tested. Experiments that made subjects orient themselves "straight ahead" found that proprioceptive input from the neck region does significantly impact the subjective body orientation in humans, even though the effect was smaller than that found for vestibular stimulation [97]. In order to clarify whether the sensorimotor information integrated with acoustic cues is derived from vestibular or proprioceptive systems, several head and and body movement conditions were tested in a sound localisation experiment [98]. The proprioceptive information did not improve localisation indicating that

the vestibular inputs are sufficient to inform the auditory system about head movement. In line with these findings, work by Genzel et al. shows that auditory updating is dominated by vestibular signals, though they did find significant contributions from proprioception/efference copy [99]. Even eye position [100] and audiovestibular interaction [101] seem to affect the spatial localisation, which further complicates the understanding of the contribution of sensorimotor information to the process of sound localisation. It is apparent, however, that the vestibular system is dominant in many of the tested scenarios.

3.3 Active-listening strategies

Borrowing from control theory terminology, active listening can be subdivided into open-loop and closed-loop listening. In open-loop listening head movements do not depend on the sound source. For closed-loop listening, the listener adjusts the head movements in response to the perceived sound in a feedback loop, adapting the movement for the duration of the sound signal. This makes closed-loop listening a task-dependent problem. Closed-loop listening can be beneficial to "triangulate" a source, to decrease interference from reverberation, and to attend to a single (moving) source in a complex listening environment [52]. Naturally, closed-loop listening strategies are only possible if there is enough time to react. In a dynamic listening task, localisation accuracy could be improved with signals as short as 50 ms, but only in the cases where the listener responds with a head movement within the duration of the stimulus [6]. Note that, besides head movement, eyes can also be moved as a reaction to a sound [102], but because they do not change the auditory signal, we do not consider them in this article.

In an unconstrained listening situation, i.e., any head movements allowed, listeners utilise yaw more often than pitch or roll [62, 103]. This is in line with the observation that yaw produces the most informative dynamic cues, compare Figure 3. In a closed-loop listening situation, listeners can also orient their head towards the source. Indeed, in another listening task, a majority of the subjects rotated their head toward the direction of the source [104, 105]. This behaviour may be beneficial for several reasons such as the spatial centring the acoustic image of a sound and the alignment of the visual system with the source of the sound [106]. The horizontal localisation is best in the area around the frontal half of the median plane [107, 108] and thus may be considered as a neuro-computational auditory fovea, which somewhat resembles the visual ocular pursuit system [74]. Following this, a listener's intention may be to orient their head such that the source direction is within the field of highest spatial resolution [3]. Furthermore, listeners also tend to make reversals in head movements, i.e., rotating their head back and forth [104, 105]. By doing so, a continuum of dynamic cues is produced, which when integrated, may improve the estimation of the position of a stationary sound source.

It may be unnecessary to consider all physically attainable orientations of the head, because a confined area around the initial position covers the majority of head positions in natural listening situations. Even though humans can rotate their heads on the yaw axis as far as $\pm 70°$, the listeners do not seem to rotate their heads to this extent [7]. Small head rotations (up to $\pm 16°$) already significantly reduce the rate of front–back confusions, though larger movements are required to also significantly reduce elevation errors [38]. Head movements are smaller for broadband noise than for narrowband noise [62], indicating an inverse relationship between spectral content and the required rotation angle.

It is important to note that all the aforementioned studies report large individual differences in the head movements. The optimal manner of obtaining dynamic information may be subject-dependent because of differences in morphology and hearing capabilities. It is, however, rather likely that untrained and fully unconstrained listeners do not inherently know how to utilise dynamic cues. In a speech perception experiment, untrained listeners did not make optimal use of the dynamic cues [109]. In fact, some listeners did not move at all, some rotated directly to near-optimum orientations, while others moved gradually and erratically. After being instructed on the head movements, listeners' behaviour became more coherent and performance improved indicating that listeners are capable of quickly learning new strategies in order to optimise their head movements. However, there seem to be only little advantage from "free" (i.e., no instructions) over "forced" (i.e., an instructed direction and speed) rotations [6, 59]. In summary, inclusion of individual listener strategies in an active listening model would require the consideration of a task-dependent variable driving the head orientation, freely chosen at each moment of time.

4 Bayesian models

There is a general consensus that in order to estimate a sound-source direction, the human auditory system performs a comparison between incoming acoustic features and their learned representation [11, 31, 34, 110]. In other words, the models assume a template-matching process, in which the auditory system maintains a stored library of templates of the acoustic information associated with each sound-source direction. When a stimulus is perceived, the listener then compares it to the templates. Given some prior assumptions, the localisation estimate then corresponds to the direction for which the template fits most closely.

This procedure can be well represented in the Bayesian framework, in which the probability of an occurring event may be affected by prior knowledge about the event, e.g., how frequently a stimulus previously occurred at a given position. The inclusion of a prior probability is what distinguishes this method from other interpretations of probability. A multitude of studies on multimodal perception suggests that the brain uses a Bayesian approach to combine stimuli during estimation of spatial localisation [16, 20, 111–113] and to learn and adapt to changes in the environment [45, 72, 114–116].

4.1 Bayesian estimation

In Bayes' theorem (in terms of probability density functions),

$$p(\psi|\mathbf{y}) = \frac{p(\mathbf{y}|\psi)p(\psi)}{p(\mathbf{y})},$$

the posterior probability density function (PDF) $p(\psi|\mathbf{y})$ of direction ψ of a source given acoustic information $\mathbf{y}$ depends on three factors: (1) The likelihood $p(\mathbf{y}|\psi)$, representing the PDF of acoustic information $\mathbf{y}$ being observed for a source at direction ψ; (2) The prior PDF $p(\psi)$, representing assumptions on the result, derived from the past experience on the parameter to be estimated; and (3) The denominator $p(\mathbf{y})$, representing the PDF of acoustic information $\mathbf{y}$ being observed and assumed to be a normalisation constant inferred from $\int p(\psi|\mathbf{y})\mathrm{d}\psi = 1$, so that the area under the posterior PDF integrates to 1.

When formulating the sound source localisation problem as a Bayesian decision problem, the listener first determines the posterior PDF given both prior and sensory information. Next, a loss function is defined on the set of source directions and by using the posterior PDF to minimise the expected loss the "best" estimate of the source direction is determined.

If the loss function specifies the minimisation of the probability of error, the optimal Bayesian decision rule selects the maximum of the posterior PDF, a strategy known as the maximum-a-posteriori (MAP) strategy [117]. Note that in the special case of the prior being a uniform PDF, the MAP strategy obtains the same result as maximum likelihood estimation (MLE), which returns the parameter value ψ that maximises the likelihood. Interestingly, the auditory system does not always rely on a point estimate like the MAP rule and a random sampling strategy from the posterior PDF seems to better explain the localisation process in some conditions [42].

Bayesian inference is a widely used approach in investigating various auditory effects. For example, Bayesian inference can help to reconstruct localisation cues from listener responses to random cue spectra [118], or to investigate how fluctuations of binaural cues in realistic noisy listening conditions affect localisation performance [119], or to investigate the trading between accuracy and precision in the sound-localisation process [42].

Bayesian inference and the template-matching procedure have been combined to model sound localisation based on ITDs and spectral acoustic features of a stationary source [11]. That Bayesian ideal-observer model was able to predict empirical sound localisation errors, reproducing patterns observed in human localisation experiments. It can be seen as a first step in modelling active dynamic localisation and is, in fact, a simplified example for our model later (see Sect. 5.5).

4.2 Recursive Bayesian estimation

There is a variety of methods to introduce time dependency to Bayesian models [120], especially when it comes to derive a decision in complex dynamic systems [121]. Recursive Bayesian inference is one of these techniques and can be applied to fit a statistical model to data in a series of steps [122]. The methodology of recursive Bayesian inference can be defined as a two-step process that recursively cycles through a prediction step (which "predicts" a prior PDF of current state based on the old estimate) and an update step (which "updates" the state PDF to form a posterior estimate by taking the newly available measurement into account).

A popular way to approximate the recursive Bayesian estimation in discrete state-space is through linear-Gaussian models, i.e., Kalman filters, which assume that the true state of system model $\mathbf{X}(t_i)$ at time t_i linearly evolves from the state at time t_{i-1}. In order to process non-linear systems, adaptations of the Kalman filter can be used, such as extended Kalman filters or unscented Kalman filters [123]. The Kalman filter and its variants update the process mean, i.e., the state, and its variance at each iteration, making it the optimum (minimum error) estimator when the noises are Gaussian. However, multimodal noises may bring the Kalman filter to instability, which prevents it from converging to the mean. In order to handle such multimodal ambiguities as well as non-linear models, the particle filter method (or sequential Monte Carlo, SMC) has been developed [124].

The switch between linear and non-linear systems can even be required when solving a single problem. For example, in an investigation of auditory-based prey capture by the barn owl [24], both linear and non-linear recursive Bayesian estimations were used to predict a source's future direction, given a sequence of sensory observations and a prior PDF for direction and angular velocity. A linear relationship between prey direction and ITD was assumed for prey in the frontal hemisphere allowing for the use of a Kalman filter. For more lateral sound directions, however, the linear approximation did not apply and a particle filter was required to compute the Bayesian prediction. Note that while that model is a dynamic model, it utilises dynamic cues from source motion only, not from head movements.

Recursive Bayesian inference was also used to model the external nucleus of the inferior colliculus [9], which is thought to be responsible for the transformation from a frequency-specific code for spatial cues into a topographic code for space [125]. A similar mechanism was able to explain how an input with a limited spectrum can improved elevation estimation after training [25].

In the field of auditory scene analysis [126], recursive Bayesian estimation has been used to improve the process of resolving individual sources in a complex acoustic environment [127]. That model is based on dynamic ITDs and uses a simple recursive approach, in which the prior PDF of sound positions is stored in a spectrospatial map and the incoming short-time maps constitute new evidence. As a simplification, the dynamic cues of the head movements are translated to equivalent inverse changes in the source position. The model also assumes perfect control of the head orientation. Thus, it does not reflect a realistic dynamic listening situation, in which the actual head orientation

needs to be a random variable with an uncertainty, because of the inherent error in the sensorimotor system. Still, it can be considered as a case of "idealised" dynamic listening.

Much can also be learnt from research in robotics, as there are many existing models for multisensory integration and motion strategies [14, 128–131]. However, many of these studies have focused on developing artificial auditory systems that are less applicable to research on human binaural listening, such as the use of microphone arrays [132, 133]. Nonetheless, the techniques applied can prove useful for more biologically plausible models, especially for dynamic listening. For example, performance of an extended Kalman filter has been improved by introducing additional a priori information verifying the consistency of location estimation at each time step [89].

5 Modelling active listening in a static environment

In this section, we propose a state-space model to describe the problem of active sound localisation, explain the generative model used in the Bayesian framework, and derive the posterior PDFs of the head orientation and source direction. We also demonstrate the feasibility of our model in two simplified examples.

In our model, we limit the source to be stationary with respect to the head movements within the considered temporal interval in the model. This is justified by assuming that our framework is a part of a larger framework of causal inference, in which the listener tests various hypotheses on the auditory environment and at the moment of probing the environment, the most probable assumption is an environment consisting of non-moving auditory objects. This assumption is further upheld by the empirical data supporting the "slow-motion prior" in listening [48, 49].

We allow the listener to actively control the head movements. While head rotations can be described in a general way by using quaternions, in our article, we use a simplified description by limiting head movements to yaw rotations only. Finally, we assume a "multiple looks" model whereby the listener collects information at discrete time steps during the motion and updates the evidence step by step.

5.1 State-space model

We describe the listener as a dynamic state-space process, where in all instances of listening, the listener needs to determine the posterior PDF of the source direction and that PDF needs to be updated recursively as more evidence or information becomes available. This recursive estimation process is important in dynamic models including temporal integration and relies on the Markov assumption: the future is independent of the past given the present [134]. The state-space representation utilises state variables, of which the values evolve over time in a way that depends on their current value (i.e., state) and on the input variables. Using this definition and the Markov assumption, we denote the true state of the system as $\mathbf{X}$. This true state is hidden to the listener, who can only observe the state via a noisy measurement that we denote as $\mathbf{y}$,

$$\begin{aligned}\mathbf{X}(t_i) &= g(\mathbf{X}(t_{i-1}), u(t_{i-1}), \delta_u(t_i)),\\ \mathbf{y}(t_i) &= h(\mathbf{X}(t_i), \delta_x(t_i)),\end{aligned} \tag{1}$$

where g and h are called the system model and the measurement model, respectively, u is the system control function, and δ_u and δ_x are the system and measurement noise, respectively.

In the context of a sound localisation process with the source positioned in the far-field, the state information required by the listener to localise a source consists of the head orientation $\theta_H(t)$ and the source direction ψ a 2D vector defined by the lateral and polar angles of the source,

$$\mathbf{X}(t) = (\theta_H(t), \psi)^T,$$

with both the head orientation and the source direction measured relative to the torso that we assume as the link between the egocentric and allocentric spatial systems.

The control signal $u(t)$ is represented by the speed of rotating the head, i.e., $u(t) = \omega_z(t)$. Thus, the discrete-time dynamic state-space model g is formulated as,

$$\mathbf{X}(t_{i+1}) - \mathbf{X}(t_i) = (\omega_z(t_i)\Delta t + \delta_u, \psi(t_{i+1}) - \psi(t_i))^T, \tag{2}$$

with Δt the time step of the "multiple looks" model and δ_u the noise on the self-motion representing the difference between the intended and the actually executed head movement. Note that, in a stationary auditory environment, the difference between the previous and current sound direction, $\psi(t_{i+1})$ and $\psi(t_i)$, respectively, is zero.

The measurement equation of the proposed state-space model consists of two components,

$$\mathbf{y}(t_i) = (y_A(t_i), y_H(t_i))^T.$$

The first component, $y_A(t_i)$, describes the acoustic features used by the auditory system in the process of sound localisation, as described in Section 3. For example, when resolving front–back confusions, the system may rely on dynamic ITDs only, resulting in $y_A(t_i) = \mathrm{ITD}(\theta_H(t_i), \psi) + \delta_A$. In that example, the noise in the acoustic features is modelled as unbiased Gaussian noise $\delta_A \sim \mathcal{N}(0, \sigma_{y_{ITD}})$.

The second component, $y_H(t_i)$, describes the measurement of the head orientation as $y_H(t_i) = \theta_H(t_i) + \delta_H$ and assumes that the listener knows the head orientation relative to the torso up to some additive (unbiased) Gaussian noise $\delta_H \sim \mathcal{N}(0, \sigma_{y_H})$.

While the aforementioned noise sources are described as additive, control of movement [135] and stimulus perception [136] are generally assumed to not only be endowed with additive but also with multiplicative noise, in which the standard deviation of the noise is linearly related to the amplitude of the signal. A possible solution in the case of our framework can be transforming the signals to a space of constant variance.

Note that we make no assumptions about the linearity of the acoustic component of the measurement equation.

Note also that we do assume that the head orientation and acoustic measurement processes are independent, in particular that the acoustic features are not used by the listener to estimate the head orientation. This is directly exploited in the following sections: in order to translate the acoustic measurements into source direction, the listener needs information about the head orientation and we simplify the active sound localisation problem considerably by first estimating the head orientation (Sect. 5.3), and subsequently making use of that information in the process of estimating the source position (Sect. 5.4). In order to retain the assumption of independent noise in the underlying measurement processes, we assume the delays involved in the process to be broadband.

5.2 Generative model

In the framework of Bayesian inference, we assume that the listener wants to determine the source direction based on all prior information about the environment and on all sensory information collected during the head movement. To this end, the listener first determines the posterior PDF of the source direction. The desired posterior PDF, taking into account all information available at time t_i will be denoted as,

$$p_{t_i} = p(\psi | u(0 \ : \ i-1), \mathbf{y}(0 \ : \ i)), \tag{3}$$

with $u(0 : i-1)$ denoting the sequence of control signals, i.e., rotation speed, applied at times t_0 until t_{i-1} with the time step Δt, and $\mathbf{y}(0 : i)$ the sequence of sensor readings, i.e., acoustic features and head orientation, collected at times t_0 until t_i. Figure 5 illustrates the generative model describing this posterior PDF. Note that, while we assume the source to be stationary, we make explicit the time varying nature of our knowledge about the source direction by taking into account all relevant information available at time t_i. Thus, we refer to this distribution by the shorthand p_{t_i}.

In Bayesian decision theory, the posterior PDF is usually used to minimise a loss function, that describes the optimal point estimate of the source direction. In closed-loop listening, the posterior PDF, computed at each update of the recursive process, can be further used as the input for a head-movement strategy. An example is the smooth posterior mean (SPM) strategy [130], which makes the listener steer the head on a smooth trajectory towards the posterior mean of the source position during each iteration. While the definition of a relevant loss function and its minimisation is both of theoretical interest and required for practical implementations (see our numerical examples in Sect. 5.5), it is beyond the scope of this article having its main focus on the derivation of the posterior PDF.

5.3 Estimation of head orientation

We model the dynamic process of the head rotation by,

$$\theta_H(t_{i+1}) = \theta_H(t_i) + \omega_z(t_i)\Delta t + \delta_u,$$
$$y_H(t_{i+1}) = \theta_H(t_{i+1}) + \delta_H,$$

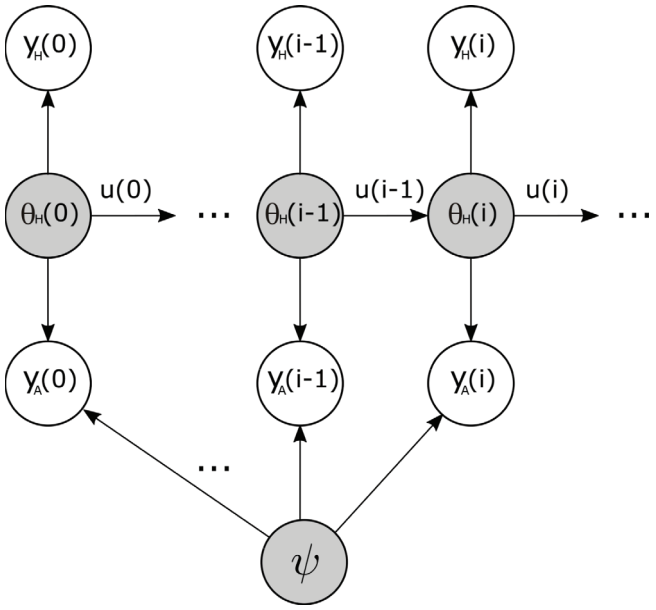

Figure 5. Bayesian network describing the dynamic listening situation. The white and grey circles represent observed and hidden variables, respectively. The arrows denote conditional dependencies. ψ denotes the stationary sound source direction.

with the initial head orientation given by a normally distributed variable $\theta_H(t_0) \sim \mathcal{N}(\theta_0, \sigma_0)$ representing the listener's uncertainty about the initial head orientation. The additive noise on both the movement equation and the sensor equation is assumed to be zero-mean white Gaussian noise $\delta_u \sim \mathcal{N}(0, \sigma_u)$ and $\delta_H \sim \mathcal{N}(0, \sigma_{y_H})$.

Making use of Bayes' rule and taking into account all head rotations executed as well as all sensor readings collected so far, the PDF of the head orientation at time t_{i+1} during the head movement can be shown to be Gaussian and given by,

$$p(\theta_H(t_{i+1}) | y_H(0 \ : \ i+1), u(0 \ : \ i)) = \mathcal{N}(\hat{\theta}_H(t_{i+1}), \sigma_{\theta_H(t_{i+1})}), \tag{4}$$

with mean and variance,

$$\hat{\theta}_H(t_{i+1}) = (1-K) \cdot (\hat{\theta}_H(t_i) + \omega_z(t_i)\Delta t) + K \cdot y_H(t_{i+1}),$$
$$\sigma^2_{\theta_H(t_{i+1})} = (1-K) \cdot (\sigma^2_{\theta_H(t_i)} + \sigma^2_u),$$

and,

$$K = \frac{\sigma^2_{\theta_H(t_i)} + \sigma^2_u}{\sigma^2_{\theta_H(t_i)} + \sigma^2_u + \sigma^2_{y_H}}.$$

The prior required to initiate the recursive process is based on two components: the prior knowledge available to the listener about the sound source direction $p(\psi)$ and the prior knowledge $p(\theta_H(0) \mid y_H(0))$ available to the listener about the initial head orientation. We assume here that $\theta_H(0)$ is a Gaussian distribution centred on an initial head orientation $\hat{\theta}_H(0)$ as described in Equation (6), but that may depend on the actual experiment being modelled.

Note the recursive nature of these equations as well as their correspondence with a Kalman filter implementation of the head orientation estimation process.

5.4 Estimation of sound-source direction

Here we derive a recursive expression for Equation (3) describing the posterior PDF p_{t_i} at time $t_i = t_{i-1} + \Delta t$ in terms of the prior PDF $p_{t_{i-1}}$ derived at time t_{i-1} combined with the extra information from the most recent "look" in the sequence of "multiple looks" collected during the head movement. We assume the sensor readings and the control signals to be available to the estimation process as,

$$\mathbf{y}(0 \ : \ i) = ((y_A(t_0), y_H(t_0))^T, (y_A(t_1), y_H(t_1))^T, \cdots (y_A(t_i), y_H(t_i))^T),$$

and,

$$u(0 \ : \ i-1) = [\omega_z(t_0), \omega_z(t_1) \cdots \omega_z(t_{i-1})],$$

i.e., we assume a varying speed of head rotation around the yaw-axis (which remains constant during each time-step Δt).

In the first step, as the source direction is part of the full state $\mathbf{X} = (\theta_H, \psi)^T$, we derive the desired PDF described by Equation (3) from the joint full-state PDF by marginalisation over all possible head orientations,

$$\begin{aligned} p_{t_i} &= p(\psi | \mathbf{y}(0 \ : \ i), u(0 \ : \ i-1)) \\ &= \int_{\theta_H} p_X(\psi, \theta_H(t_i) \,|\, \mathbf{y}(0 \ : \ i), u(0 \ : \ i-1)) \mathrm{d}\theta_H. \end{aligned}$$

This operation allows us to correctly take into account the effect of the remaining head orientation uncertainty on the source estimation. The joint PDF p_X can be expanded as,

$$\begin{aligned} &p_X(\psi, \theta_H(t_i) | (y_A(t_i), y_H(t_i))^T, \mathbf{y}(0 \ : \ i-1), u(0 \ : \ i-1)) \\ &\quad = p(\psi | \theta_H(t_i), y_A(t_i), \mathbf{y}(0 \ : \ i-1), u(0 \ : \ i-1)) \\ &\quad \times p(\theta_H(t_i) | y_H(t_i), \mathbf{y}(0 \ : \ i-1), u(0 \ : \ i-1)). \end{aligned}$$

Our knowledge of the head orientation taking into account all relevant data up until time t_i can be described by a reformulation of Equation (4),

$$p(\theta_H(t_i) | y_H(0 \ : \ i), u(0 \ : \ i-1)) = \mathcal{N}(\hat{\theta}_H(t_i), \sigma_{\theta_H(t_i)}).$$

In the second step, by using Bayes' rule, we rewrite the first term of p_X,

$$\begin{aligned} &p(\psi | \theta_H(t_i), y_A(t_i), \mathbf{y}(0 \ : \ i-1), u(0 \ : \ i-1)) \\ &= \frac{p(y_A(t_i) | \theta_H(t_i), \psi) \times p(\psi | \mathbf{y}(0 \ : \ i-1), u(0 \ : \ i-1))}{p(y_A(t_i) | \theta_H(t_i), \mathbf{y}(0 \ : \ i-1), u(0 \ : \ i-1))}, \end{aligned}$$

and simplify it with,

$$p(\psi | \mathbf{y}(0 \ : \ i-1), u(0 \ : \ i-1)) = p_{t_{i-1}},$$

i.e., the posterior distribution on the source direction we determined at time t_{i-1}.

By combining these results we obtain the recursive expression for the posterior PDF,

$$\begin{aligned} &p(\psi | \mathbf{y}(0 \ : \ i), u(0 \ : \ i-1)) \\ &= p_{t_i} = C \cdot p_{t_{i-1}} \times \int_{\theta_H} p(\theta_H(t_i) \,|\, y_H(0 \ : \ i), u(0 \ : \ i-1)) \\ &\qquad \times p(y_A(t_i) \,|\, \theta_H(t_i), \psi) \mathrm{d}\theta_H, \end{aligned} \tag{5}$$

with the normalisation constant C derived from the posterior PDF,

$$\int_{\psi} p(\psi | \mathbf{y}(0 \ : \ i), u(0 \ : \ i-1)) \mathrm{d}\psi = 1.$$

The recursive process in Equation (5) is comparable to [127] and expresses our new state of knowledge about the source direction p_{t_i} based on the previous state of knowledge $p_{t_{i-1}}$ with the extra knowledge obtained in the most recent "look". This extra knowledge takes into account not only the acoustic measurement but also the current head orientation estimate.

The recursive processes is initiated with the source direction PDF from Equation (3) derived from the initial acoustic measurement performed at time t_0,

$$p_{t_0} = p_\psi(\psi | (y_A(t_0), y_H(t_0))^T).$$

This PDF, following a similar derivation as the one described for Equation (5), is given by,

$$\begin{aligned} p_{t_0} &= C \cdot p(\psi) \times \int_{\theta_H} p(\theta_H(t_0) | y_H(t_0)) \\ &\qquad \times p(y_A(t_0) | \theta_H(t_0), \psi) \mathrm{d}\theta_H, \end{aligned} \tag{6}$$

with $p(\theta_H(t_0)|y_H(t_0)) = \mathcal{N}(\hat{\theta}_H(t_0), \sigma_{\theta_H(t_0)})$ representing the initial (uncertain) head orientation and $p(\psi)$ being the prior PDF on the source direction. Note that at this very first moment t_0 the head motion is not made use of, as at that moment only the current head orientation is known (up to some uncertainty). As with Equation (5), the constant C can be derived from the normalisation of this posterior PDF.

Depending on the behavioural task and listener's priors on the environment, this prior information $p(\psi)$ (in Eq. (6)) may substantially modify the model predictions. For example, it may constrain the possible source directions to a sub-region of the full sphere around the listener's head, e.g., the frontal hemisphere or the horizontal plane. This can be modelled by choosing $p(\psi)$ accordingly.

5.5 Numerical examples

The presented concept is mathematically consistent, yet its complete numerical evaluation is not trivial: it depends on the considered acoustic features and needs to consider many listening situations. Thus, it deserves separate discussions in future articles. In this article, we illustrate the

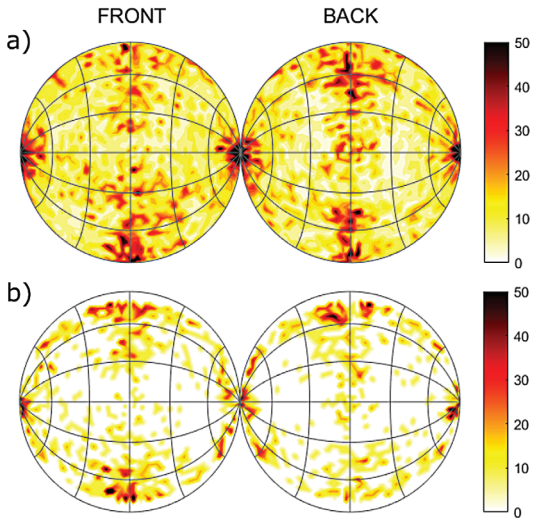

Figure 6. Example 1: Predictions obtained from 10 simulated trials of the simplified concept without head movements, head orientation straight ahead. (a) Polar angle errors (in degrees). (b) Front–back confusion rates (in %); rates for target directions near the frontal plane are not shown for clarity.

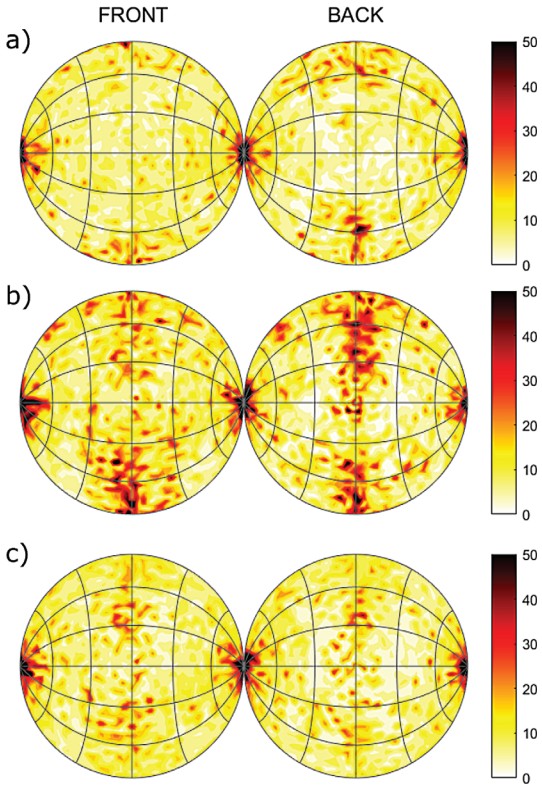

Figure 7. Example 2: Polar angle errors (in degrees) obtained from 10 simulated trials of the simplified concept with a 10° turn and perfectly determinable head orientation. (a) Head yaw rotations. (b) Head pitch turns. (c) Head roll tilts. Left and right panels represent source locations in the front and back of the head, respectively.

explanatory power of the concept by numerically applying it to two examples in a simplified setting, using an MAP estimator to convert the posterior PDF into a point estimate. The corresponding code was implemented in the auditory modeling toolbox (AMT) version 1.0 (Majdak et al. [137]). Note that the MAP estimator is just one of the possibilities to obtain a point estimate from the PDF.

In the first example, we simplify our concept to the sound localisation with head oriented straight ahead and without any head movements (not even noise), i.e., $u = 0$, and $y_H = 0$, respectively. This simplifies Equation (2) to $\mathbf{X}(t_{i+1}) = \mathbf{X}(t_i)$ and Equation (3) to $p = p_\psi(\psi|y_A)$. The resulting model corresponds to the ideal-observer model by Reijniers et al. [11]. Results from that model were replicated by using our implementation of the simplified concept. To replicate the results from the original study, we used the same acoustic features y_A: ITDs, the summed binaural spectral information, and the interaural spectral information. We also used the same noise parameters tuned to the ITD thresholds and absolute hearing thresholds. For the input signal, we used Gaussian white noise filtered with HRTFs from Section 3.1, which were also used as templates in the model. 2354 target directions on the spherical grid were considered. Per target sound direction, 10 trials were simulated and averaged to obtain polar errors and front–back confusion rates. Figure 6 shows the polar errors and front–back confusion rates obtained with that simplified model. The polar errors are largest above and below the listener, which is qualitatively in line with the observations obtained from actual localisation experiments [138]. The large polar errors near the interaural axis can be attributed to the disproportional changes in polar angles even for small changes of the source direction [26].

The second example demonstrates the recursive nature of the estimation process in our concept. We reduce the concept to head rotations that are small (i.e., up to 10°), open-loop, and at a constant speed (i.e., $u = \omega_z$), so we can assume a linear relationship between ITD and head-rotation angle [24]. Head position is assumed to be exactly known, i.e., a deterministic y_H not confounded by any noise. Further, we used the broadband ITD as acoustic feature, i.e., $y_A(t_i) = \mathrm{ITD}(\theta_H(t_i), \psi) + \delta_A$ and used a uniform distribution for the prior PDF. Taking into account these simplifications, Equation (3) reduces to $p_{t_i} = p(\psi \mid \mathbf{y}(0 : i))$ with $\mathbf{y}(t_i) = (y_A(t_i), \theta_H(t_i))$. The predictions were calculated for yaw, roll, and pitch separately, with identical simulation parameters from the first example. Figures 7 and 8 show the polar errors and front–back confusion rates, respectively. The results show that head yaw significantly reduces front–back confusions, though roll also shows a slight improvement. Head yaw and roll also reduced polar errors, with most of the reduction for sources located at the eye level. All these findings are in line with empirical data [6, 38]. Head pitch, on the other hand, did not show much improvement, neither in reducing the localisation errors nor the front–back confusions, both being in line with the observation of small ITD rates when tilting the head (see Fig. 3c).

These two examples demonstrate the feasibility of the concept.

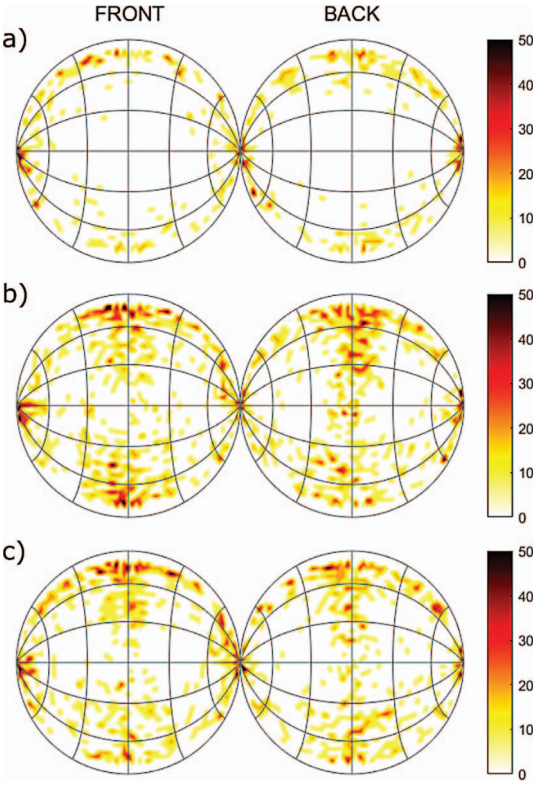

Figure 8. Example 2: Front–back confusion rates (in %) obtained from the simplified concept with a 10° turn and perfectly determinable head orientation. (a) Head yaw rotations. (b) Head pitch turns. (c) Head roll tilts. For clarity rates for target directions near the frontal plane are not shown. Left and right panels represent source locations in the front and back of the head, respectively.

6 Conclusions

This article briefly reviews the recent literature on modelling active dynamic sound localisation, which complements the well-documented sound localisation based on static acoustic features with cues from self-motion or source motion. The review focuses on Bayesian inference because of its prominent role in recent multimodal cognitive models and high potential in modelling dynamic cognitive processes. We have defined the term of active dynamic sound localisation, describing the localisation process in which the listener actively updates the head orientation to facilitate the localisation process.

Further, we described a theoretical Bayesian modelling framework based on the independent estimation of acoustic features and head rotations. In order to show the feasibility of the concept, we provide two short examples of simplified versions of the concept, for which numeric implementations are available. While these two examples do not fully validate the concept in all its aspects, they demonstrate the potential of the proposed concept towards a general dynamic sound localisation model.

Future work will involve model extensions along different directions. First, model parameters will need to be fine-tuned through sensitivity analyses and comparisons to empirical data in order to quantitatively fit the predictions to the sound localisation performance of humans. Second, an implementation of a closed-loop version of the model will be required to completely test the concept. Here questions related to listening strategies will become relevant. Additionally, our current concept considers stationary sources only. It can be extended to dynamic auditory environments by integrating e.g., a multiscale network [139], expanding our concept to a general framework of active sound localisation in dynamic auditory environments.

Conflict of interest

Author declared no conflict of interests.

Acknowledgments

This research was supported by the Research Foundation Flanders (FWO) under Grant no. G023619N, the Agency for Innovation and Entrepreneurship (VLAIO), and the European Union (EU, project "SONICOM", grant number 101017743, RIA action of Horizon 2020).

Data availability statement

Implementations of both the model (mclachlan2021) and the simulations (exp_mclachlan2021) are publicly available as part of the Auditory Modeling Toolbox (AMT, https://www.amtoolbox.org) [137] in the release of the version 1.0.0 available as a full package for download [140].

References

1. P. Avan, F. Giraudet, B. Büki: Importance of binaural hearing. Audiology and Neurotology 20, Suppl. 1 (2015) 3–6.
2. J. Blauert, J. Braasch, eds.: The technology of binaural understanding, Modern acoustics and signal processing. Springer International Publishing, 2020. https://www.springer.com/gp/book/9783030003852.
3. J. Blauert: Spatial hearing: The psychophysics of human sound localization. MIT Press, 1997.
4. J. Tobias: Foundations of modern auditory theory. Elsevier, 2012.
5. F.L. Wightman, D.J. Kistler: Monaural sound localization revisited. The Journal of the Acoustical Society of America 101, 2 (1997) 1050–1063.
6. S. Perrett, W. Noble: The contribution of head motion cues to localization of low-pass noise. Perception & Psychophysics 59, 7 (1997) 1018–1026.
7. C. Kim, R. Mason, T. Brookes: Head movements made by listeners in experimental and real-life listening activities. Journal of the Audio Engineering Society 61 (2013) 425–438.
8. E.A. Macpherson: A computer model of binaural localization for stereo imaging measurement. Journal of the Audio Engineering Society 39, 9 (1991) 604–622.
9. V. Willert, J. Eggert, J. Adamy, R. Stahl, E. Korner: A probabilistic model for binaural sound localization. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 36, 5 (2006) 982–994.

10. R. Baumgartner, P. Majdak, B. Laback: Modeling sound-source localization in sagittal planes for human listeners. The Journal of the Acoustical Society of America 136, 2 (2014) 791–802.
11. J. Reijniers, D. Vanderelst, C. Jin, S. Carlile, H. Peremans: An ideal-observer model of human sound localization. Biological Cybernetics 108, 2 (2014) 169–181.
12. J. Braasch: Localization in the presence of a distracter and reverberation in the frontal horizontal plane: II. Model algorithms. Acta Acustica United with Acustica 88, 6 (2002) 956–969.
13. T. May, S. Van De Par, A. Kohlrausch: A probabilistic model for robust localization based on a binaural auditory front-end". IEEE Transactions on Audio, Speech, and Language Processing 19, 1 (2010) 1–13.
14. N. Ma, T. May, G.J. Brown: Exploiting deep neural networks and head movements for robust binaural localization of multiple sources in reverberant environments. IEEE/ ACM Transactions on Audio, Speech, and Language Processing 25, 12 (2017) 2444–2453.
15. A. Kothig, M. Ilievski, L. Grasse, F. Rea, M. Tata: A bayesian system for noise-robust binaural sound localisation for humanoid robots, in 2019 IEEE International Symposium on Robotic and Sensors Environments (ROSE), IEEE. 2019, pp. 1–7.
16. D. Alais, D. Burr: The ventriloquist effect results from near-optimal bimodal integration. Current Biology 14, 3 (2004) 257–262.
17. P.W. Battaglia, R.A. Jacobs, R.N. Aslin: Bayesian integration of visual and auditory signals for spatial localization. The Journal of the Optical Society of America A 20, 7 (2003) 1391–1397.
18. M.O. Ernst, M.S. Banks: Humans integrate visual and haptic information in a statistically optimal fashion. Nature 415, 6870 (2002) 429–433.
19. D.C. Knill, A. Pouget: The bayesian brain: The role of uncertainty in neural coding and computation. TRENDS in Neurosciences 27, 12 (2004) 712–719.
20. L. Shams, W.J. Ma, U. Beierholm: Sound-induced flash illusion as an optimal percept. Neuroreport 16, 17 (2005) 1923–1927.
21. R.A. Jacobs: Optimal integration of texture and motion cues to depth. Vision Research 39, 21 (1999) 3621–3629.
22. H.H. Bülthoff, H.A. Mallot: Integration of stereo, shading and texture, in 11th European Conference on Visual Perception (ECVP 1988), Wiley. 1990, pp. 119–146.
23. M.S. Landy, L.T. Maloney, E.B. Johnston, M. Young: Measurement and modeling of depth cue combination: in defense of weak fusion. Vision Research 35, 3 (1995) 389–412.
24. W. Cox, B.J. Fischer: Optimal prediction of moving sound source direction in the owl. PLoS Computational Biology 11, 7 (2015) e1004360.
25. B. Zonooz, E. Arani, A.J. Van Opstal: Learning to localise weakly-informative sound spectra with and without feedback. Scientific Reports 8, 1 (2018) 1–14.
26. P. Majdak, M.J. Goupell, B. Laback: 3-d localization of virtual sound sources: Effects of visual environment, pointing method, and training. Attention, Perception, & Psychophysics 72, 2 (2010) 454–469.
27. R. Barumerli, P. Majdak, J. Reijniers, R. Baumgartner, M. Geronazzo, F. Avanzini: Predicting directional sound-localization of human listeners in both horizontal and vertical dimensions, in Audio Engineering Society Convention 148, Audio Engineering Society. 2020.
28. E.A. Shaw: Transformation of sound pressure level from the free field to the eardrum in the horizontal plane. The Journal of the Acoustical Society of America 56, 6 (1974) 1848–1861.
29. L. Rayleigh: Xii. on our perception of sound direction. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science 13, 74 (1907) 214–232.
30. E.A. Macpherson, J.C. Middlebrooks: Listener weighting of cues for lateral angle: The duplex theory of sound localization revisited. The Journal of the Acoustical Society of America 111, 5 (2002) 2219–2236.
31. J.C. Middlebrooks: Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency. The Journal of the Acoustical Society of America 106, 3 (1999) 1493–1510.
32. M. Morimoto, H. Aokata: Localization cues of sound sources in the upper hemisphere. Journal of the Acoustical Society of Japan (E) 5, 3 (1984) 165–173.
33. R.B. King, S.R. Oldfield: The impact of signal bandwidth on auditory localization: Implications for the design of three-dimensional audio displays. Human Factors 39, 2 (1997) 287–295.
34. B. Zonooz, E. Arani, K.P. Körding, P.R. Aalbers, T. Celikel, A.J. Van Opstal: Spectral weighting underlies perceived sound elevation. Scientific Reports 9, 1 (2019) 1–12.
35. J. Hebrank, D. Wright: Spectral cues used in the localization of sound sources on the median plane. The Journal of the Acoustical Society of America 56, 6 (1974) 1829–1834.
36. J. Jiang, B. Xie, H. Mai, L. Liu, K. Yi, C. Zhang: The role of dynamic cue in auditory vertical localisation. Applied Acoustics 146 (2019) 398–408.
37. E.M. Wenzel, M. Arruda, D.J. Kistler, F.L. Wightman: Localization using nonindividualized head-related transfer functions. The Journal of the Acoustical Society of America 94, 1 (1993) 111–123.
38. K.I. McAnally, R.L. Martin: Sound localization with head movement: Implications for 3-d audio displays. Frontiers in Neuroscience 8 (2014) 210.
39. P. Zahorik, D.S. Brungart, A.W. Bronkhorst: Auditory distance perception in humans: A summary of past and present research. ACTA Acustica United with Acustica 91, 3 (2005) 409–420.
40. B.G. Shinn-Cunningham, S. Santarelli, N. Kopco: Tori of confusion: Binaural localization cues for sources within reach of a listener. The Journal of the Acoustical Society of America 107, 3 (2000) 1627–1636.
41. D. Genzel, M. Schutte, W.O. Brimijoin, P.R. MacNeilage, L. Wiegrebe: Psychophysical evidence for auditory motion parallax. Proceedings of the National Academy of Sciences 115, 16 (2018) 4264–4269.
42. R. Ege, A.J. Van Opstal, M.M. Van Wanrooij: Accuracy-precision trade-off in human sound localisation. Scientific Reports 8, 1 (2018) 1–12.
43. B.J. Fischer, J.L. Peña: Owl's behavior and neural representation predicted by bayesian inference. Nature Neuroscience 14, 8 (2011) 1061–1066.
44. C.V. Parise, K. Knorre, M.O. Ernst: Natural auditory scene statistics shapes human spatial hearing. Proceedings of the National Academy of Sciences 111, 16 (2014) 6104–6108.
45. R. Ege, A.J. Van Opstal, M.M. Van Wanrooij: A.W. Mills: On the minimum audible angle. The Journal of the Acoustical Society of America 30, 4 (1958) 237-246; S.R. Oldfield, S.P. Parker: Acuity of sound localisation: a topography of auditory space. i. normal hearing conditions. Perception 13, 5 (1984) 581-600. Eneuro 6, 2 (2019).
46. U. Beierholm, S. Quartz, L. Shams: Bayesian priors are encoded independently from likelihoods in human multisensory perception. Journal of Vision 9 (2009) 23.
47. Y. Weiss, E.P. Simoncelli, E.H. Adelson: Motion illusions as optimal percepts. Nature Neuroscience 5, 6 (2002) 598–604.

48. I. Senna, C.V. Parise, M.O. Ernst: Hearing in slow-motion: Humans underestimate the speed of moving sounds. Scientific Reports 5, 1 (2015) 1–5.
49. T.C. Freeman, J.F. Culling, M.A. Akeroyd, W.O. Brimijoin: Auditory compensation for head rotation is incomplete. Journal of Experimental Psychology: Human Perception and Performance 43, 2 (2017) 371.
50. S. Carlile, J. Leung: The perception of auditory motion. Trends in Hearing 20 (2016) 2331216516644254.
51. M. Barnett-Cowan, L.R. Harris: Temporal processing of active and passive head movement. Experimental Brain Research 214, 1 (2011) 27–35.
52. M. Cooke, Y.-C. Lu, Y. Lu, R. Horaud: Active hearing, active speaking, in ISAAR 2007-International Symposium on Auditory and Audiological Research. 2007, pp. 33–46.
53. K. van der Heijden, J.P. Rauschecker, E. Formisano, G. Valente, B. de Gelder: Active sound localization sharpens spatial tuning in human primary auditory cortex. Journal of Neuroscience 38, 40 (2018) 8574–8587.
54. A. Portello, G. Bustamante, P. Danès, J. Piat, J. Manhes: Active localization of an intermittent sound source from a moving binaural sensor, in European Acoustics Association Forum Acusticum. 2014, 12 p.
55. Y.-C. Lu, M. Cooke: Motion strategies for binaural localisation of speech sources in azimuth and distance by artificial listeners. Speech Communication 53, 5 (2011) 622–642.
56. H. Wallach: The role of head movements and vestibular and visual cues in sound localization. Journal of Experimental Psychology 27, 4 (1940) 339.
57. E.A. Macpherson: Cue weighting and vestibular mediation of temporal dynamics in sound localization via head rotation, in Proceedings of Meetings on Acoustics ICA2013, Vol. 19, Acoustical Society of America. 2013, 050131 p.
58. F.L. Wightman, D.J. Kistler: Resolution of front–back ambiguity in spatial hearing by listener and source movement. The Journal of the Acoustical Society of America 105, 5 (1999) 2841–2853.
59. W.R. Thurlow, P.S. Runge: Effect of induced head movements on localization of direction of sounds. The Journal of the Acoustical Society of America 42, 2 (1967) 480–488.
60. D.R. Begault, E.M. Wenzel, M.R. Anderson: Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source, Journal of the Audio Engineering Society 49, 10 (2001) 904–916.
61. T. Ashby, T. Brookes, R. Mason: Towards a head-movement-aware spatial localisation model: Elevation, in 21st International Congress on Sound and Vibration 2014, ICSV 2014, Vol. 4. 2014, pp. 2808–2815.
62. D. Morikawa, Y. Toyoda, T. Hirahara: Head movement during horizontal and median sound localization experiments in which head-rotation is allowed, in Proceedings of Meetings on Acoustics ICA2013, Vol. 19, Acoustical Society of America. 2013, 050141 p.
63. J. Burger: Front-back discrimination of the hearing systems. Acta Acustica United with Acustica 8, 5 (1958) 301–302.
64. R. Pavão, E.S. Sussman, B.J. Fischer, J.L. Peña: Natural itd statistics predict human auditory spatial perception. eLife 9 (2020) e51927. https://doi.org/10.7554/eLife.51927.
65. B. Bernschütz: Spherical Far-Field HRIR Compilation of the Neumann KU100. Zenodo, 2020. https://doi.org/10.5281/zenodo.3928297.
66. T. Hirahara, D. Kojima, D. Morikawa, P. Mokhtari: The effect of head rotation on monaural sound-image localization in the horizontal plane. Applied Acoustics 178 (2021) 108008. https://www.sciencedirect.com/science/article/pii/S0003682X21001018.
67. J. Leung, D. Alais, S. Carlile: Compression of auditory space during rapid head turns. Proceedings of the National Academy of Sciences 105, 17 (2008) 6492–6497.
68. A. Honda, K. Ohba, Y. Iwaya, Y. Suzuki: Detection of sound image movement during horizontal head rotation. i-Perception 7, 5 (2016) 2041669516669614.
69. J. Cooper, S. Carlile, D. Alais: Distortions of auditory space during rapid head turns. Experimental Brain Research 191, 2 (2008) 209–219.
70. G.M. Gerken, V.K. Bhat, M. Hutchison-Clutter: Auditory temporal integration and the power function model. The Journal of the Acoustical Society of America 88, 2 (1990) 767–778.
71. S. Carlile, V. Best: Discrimination of sound source velocity in human listeners. The Journal of the Acoustical Society of America 111, 2 (2002) 1026–1035.
72. S. Carlile, K. Balachandar, H. Kelly: Accommodating to new ears: the effects of sensory and sensory-motor feedback. The Journal of the Acoustical Society of America 135, 4 (2014) 2002–2011.
73. T.C. Freeman, J. Leung, E. Wufong, E. Orchard-Mills, S. Carlile, D. Alais: Discrimination contours for moving sounds reveal duration and distance cues dominate auditory speed perception. PLoS One 9, 7 (2014) e102864.
74. J.A.G.-U. Calvo, M.M. van Wanrooij, A.J. Van Opstal: Adaptive response behavior in the pursuit of unpredictably moving sounds. Eneuro 8, 3 (2021).
75. Y.A. Al'tman, I. Kudryavtseva, E. Radionova: The pattern of response of the inferior colliculus of the cat during the movement of a sound source. Neuroscience and Behavioral Physiology 15, 4 (1985) 318–324.
76. G.D. Pollak: Circuits for processing dynamic interaural intensity disparities in the inferior colliculus. Hearing Research 288, 1–2 (2012) 47–57.
77. N.J. Ingham, H.C. Hart, D. McAlpine: Spatial receptive fields of inferior colliculus neurons to auditory apparent motion in free field. Journal of Neurophysiology 85, 1 (2001) 23–33.
78. H. Wagner, T. Takahashi: Influence of temporal cues on acoustic motion-direction sensitivity of auditory neurons in the owl. Journal of Neurophysiology 68, 6 (1992) 2063–2076.
79. D. McAlpine, D. Jiang, T.M. Shackleton, A.R. Palmer: Responses of neurons in the inferior colliculus to dynamic interaural phase cues: evidence for a mechanism of binaural adaptation. Journal of Neurophysiology 83, 3 (2000) 1356–1365.
80. L. Boucher, A. Lee, Y.E. Cohen, H.C. Hughes: Ocular tracking as a measure of auditory motion perception. Journal of Physiology-Paris 98, 1–3 (2004) 235–248.
81. J. Kreitewolf, J. Lewald, S. Getzmann: Effect of attention on cortical processing of sound motion: An eeg study. NeuroImage 54, 3 (2011) 2340–2349.
82. J.C. Middlebrooks: Sound localization. Handbook of Clinical Neurology 129 (2015) 99–116.
83. N. Loveless, S. Levänen, V. Jousmäki, M. Sams, R. Hari: Temporal integration in auditory sensory memory: Neuromagnetic evidence. Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section 100, 3 (1996) 220–228.
84. X. Teng, X. Tian, D. Poeppel: Testing multi-scale processing in the auditory system. Scientific Reports 6, 1 (2016) 34390. https://www.nature.com/articles/srep34390.
85. N.F. Viemeister, G.H. Wakefield: Temporal integration and multiple looks. The Journal of the Acoustical Society of America 90, 2 (1991) 858–865.

86. P.M. Hofman, A.J. Van Opstal: Spectro-temporal factors in two-dimensional human sound localization. The Journal of the Acoustical Society of America 103, 5 (1998) 2634–2648.
87. J. Vliegen, T.J. Van Grootel, A.J. Van Opstal: Dynamic sound localization during rapid eye-head gaze shifts. Journal of Neuroscience 24, 42 (2004) 9291–9302.
88. C. Baumann, C. Rogers, F. Massen: Dynamic binaural sound localization based on variations of interaural time delays and system rotations. The Journal of the Acoustical Society of America 138, 2 (2015) 635–650.
89. M. Kumon, S. Uozumi: Binaural localization for a mobile sound source. Journal of Biomechanical Science and Engineering 6, 1 (2011) 26–39.
90. R.A. Lutfi, W. Wang: Correlational analysis of acoustic cues for the discrimination of auditory motion. The Journal of the Acoustical Society of America 106, 2 (1999) 919–928.
91. E. Schechtman, T. Shrem, L.Y. Deouell: Spatial localization of auditory stimuli in human auditory cortex is based on both head-independent and head-centered coordinate systems. Journal of Neuroscience 32, 39 (2012) 13501–13509. http://www.jneurosci.org/content/32/39/13501. https://doi.org/10.1523/JNEUROSCI.1315-12.2012.
92. J. Lewald, H.-O. Karnath: Vestibular influence on human auditory space perception. Journal of Neurophysiology 84, 2 (2000) 1107–1111.
93. I. Viaud-Delmon, O. Warusfel: From ear to body: The auditory-motor loop in spatial cognition. Frontiers in Neuroscience 8 (2014) 283. https://www.frontiersin.org/articles/10.3389/fnins.2014.00283/full. https://doi.org/10.3389/fnins.2014.00283.
94. W.A. Yost, X. Zhong, A. Najam: Judging sound rotation when listeners and sounds rotate: Sound source localization is a multisystem process. The Journal of the Acoustical Society of America 138, 5 (2015) 3293–3310. https://asa.scitation.org/doi/10.1121/1.4935091. https://doi.org/10.1121/1.4935091.
95. H. Goossens, A. Van Opstal: Influence of head position on the spatial representation of acoustic targets. Journal of Neurophysiology 81, 6 (1999) 2720–2736.
96. W.O. Brimijoin, M.A. Akeroyd: The moving minimum audible angle is smaller during self motion than during source motion. Frontiers in Neuroscience 8 (2014) 273.
97. H.-O. Karnath, D. Sievering, M. Fetter: The interactive contribution of neck muscle proprioception and vestibular stimulation to subjective "straight ahead" orientation in man. Experimental Brain Research 101, 1 (1994) 140–146.
98. J. Kim, M. Barnett-Cowan, E.A. Macpherson: Integration of auditory input with vestibular and neck proprioceptive information in the interpretation of dynamic sound localization cues, in Proceedings of Meetings on Acoustics ICA2013, Vol. 19, Acoustical Society of America. 2013, 050142 p.
99. D. Genzel, U. Firzlaff, L. Wiegrebe, P.R. MacNeilage: Dependence of auditory spatial updating on vestibular, proprioceptive, and efference copy signals. Journal of Neurophysiology 116, 2 (2016) 765–775.
100. J. Lewald, W.H. Ehrenstein: The effect of eye position on auditory lateralization. Experimental Brain Research 108, 3 (1996) 473–485.
101. D.C. Van Barneveld, A. John Van Opstal: Eye position determines audiovestibular integration during whole-body rotation. European Journal of Neuroscience 31, 5 (2010) 920–930.
102. H.H. Goossens, A.J. Van Opstal: Human eye-head coordination in two dimensions under different sensorimotor conditions. Experimental Brain Research 114, 3 (1997) 542–560.
103. W.R. Thurlow, J.W. Mangels, P.S. Runge: Head movements during sound localization. The Journal of the Acoustical society of America 42, 2 (1967) 489–493.
104. D. Muir, J. Field: Newborn infants orient to sounds. Child Development 50 (1979) 431–436.
105. J.H. Fuller: Head movement propensity. Experimental Brain Research 92, 1 (1992) 152–164.
106. W.O. Brimijoin, D. McShefferty, M.A. Akeroyd: Auditory and visual orienting responses in listeners with and without hearing-impairment. The Journal of the Acoustical Society of America 127, 6 (2010) 3678–3688.
107. A.W. Mills: On the minimum audible angle. The Journal of the Acoustical Society of America 30, 4 (1958) 237–246.
108. S.R. Oldfield, S.P. Parker: Acuity of sound localisation: A topography of auditory space. I. Normal hearing conditions. Perception 13, 5 (1984) 581–600.
109. J.A. Grange, J.F. Culling: The benefit of head orientation to speech intelligibility in noise. The Journal of the Acoustical Society of America 139, 2 (2016) 703–712.
110. J.C. Middlebrooks: Narrow-band sound localization related to external ear acoustics. The Journal of the Acoustical Society of America 92, 5 (1992) 2607–2624.
111. K.P. Körding, U. Beierholm, W.J. Ma, S. Quartz, J.B. Tenenbaum, L. Shams: Causal inference in multisensory perception. PLoS One 2, 9 (2007) e943.
112. Y. Gu, D.E. Angelaki, G.C. DeAngelis: Neural correlates of multisensory cue integration in macaque MSTd. Nature Neuroscience 11, 10 (2008) 1201–1210.
113. M. Ursino, A. Crisafulli, G. Di Pellegrino, E. Magosso, C. Cuppini: Development of a bayesian estimator for audio-visual integration: a neurocomputational study. Frontiers in Computational Neuroscience 11 (2017) 89.
114. K.P. Körding, D.M. Wolpert: Bayesian integration in sensorimotor learning. Nature 427, 6971 (2004) 244–247.
115. A.A. Stocker, E.P. Simoncelli: Noise characteristics and prior expectations in human visual speed perception. Nature Neuroscience 9, 4 (2006) 578–585.
116. T.E. Hudson, L.T. Maloney, M.S. Landy: Movement planning with probabilistic target information. Journal of Neurophysiology 98, 5 (2007) 3034–3046.
117. L. Bahl, J. Cocke, F. Jelinek, J. Raviv: Optimal decoding of linear codes for minimizing symbol error rate (corresp.). IEEE Transactions on Information Theory 20, 2 (1974) 284–287.
118. P.M. Hofman, A.J. Van Opstal: Bayesian reconstruction of sound localization cues from responses to random spectra. Biological Cybernetics 86, 4 (2002) 305–316.
119. J. Nix, V. Hohmann: Sound source localization in real sound fields based on empirical statistics of interaural parameters. The Journal of the Acoustical Society of America 119, 1 (2006) 463–479.
120. D. Barber, A.T. Cemgil, S. Chiappa: Bayesian time series models. Cambridge University Press, 2011.
121. C. Mark, C. Metzner, L. Lautscham, P.L. Strissel, R. Strick, B. Fabry: Bayesian model selection for complex dynamic systems. Nature Communications 9, 1 (2018) 1803. https://www.nature.com/articles/s41467-018-04241-5. https://doi.org/10.1038/s41467-018-04241-5.
122. S. Särkkä: Bayesian filtering and smoothing, Institute of Mathematical Statistics Textbooks. Cambridge University Press, Cambridge, 2013. https://www.cambridge.org/core/books/bayesian-filtering-and-smoothing/C372FB31C5D9A100F8476C1B23721A67.
123. E.A. Wan, R. Van Der Merwe, S. Haykin: The unscented kalman filter. Kalman Filtering and Neural Networks 5, 2007 (2001) 221–280.
124. H. Li: A Brief Tutorial On Recursive Estimation: Examples From Intelligent Vehicle Applications. 2014. ffhal-01011733v2f.

125. Y.E. Cohen, E.I. Knudsen: Maps versus clusters: Different representations of auditory space in the midbrain and forebrain. Trends in Neurosciences 22, 3 (1999) 128–135.
126. A.S. Bregman; Auditory scene analysis: The perceptual organization of sound. MIT Press, 1994.
127. D.A. Hambrook, M. Ilievski, M. Mosadeghzad, M. Tata: A bayesian computational basis for auditory selective attention using head rotation and the interaural time-difference cue. PLoS One 12, 10 (2017) e0186104.
128. R.C. Luo, C.-C. Chang: Multisensor fusion and integration: A review on approaches and its applications in mechatronics. IEEE Transactions on Industrial Informatics 8, 1 (2011) 49–60.
129. C. Schymura, T. Walther, D. Kolossa, N. Ma, G.J. Brown: Binaural sound source localisation using a Bayesian-network-based blackboard system and hypothesis-driven feedback, in Fourm Acusticum, European Acoustics Association. 2014.
130. C. Schymura, F. Winter, D. Kolossa, S. Spors: Binaural sound source localisation and tracking using a dynamic spherical head model, in Sixteenth Annual Conference of the International Speech Communication Association. 2015.
131. T. May, N. Ma, G.J. Brown: Robust localisation of multiple speakers exploiting head movements and multi-conditional training of binaural cues, in 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE. 2015, pp. 2679–2683.
132. P. Aarabi: The fusion of distributed microphone arrays for sound localization. EURASIP Journal on Advances in Signal Processing 2003, 4 (2003) 1–10.
133. J.-M. Valin, F. Michaud, J. Rouat: Robust localization and tracking of simultaneous moving sound sources using beamforming and particle filtering. Robotics and Autonomous Systems 55, 3 (2007) 216–228.
134. E. Fosler-Lussier: Markov models and hidden markov models: A brief tutorial. International Computer Science Institute, 1998.
135. E. Todorov: Stochastic optimal control and estimation methods adapted to the noise characteristics of the sensorimotor system. Neural Computation 17, 5 (2005) 1084–1108.
136. M.K. Stern, J.H. Johnson: Just noticeable difference, in The Corsini Encyclopedia of Psychology, John Wiley & Sons, Inc, Hoboken, NJ, USA. 2010, pp. 1–2.
137. P. Majdak, C. Hollomey, R. Baumgartner: AMT 1.x: A toolbox for reproducible research in auditory modeling. Submitted to Acta Acustica.
138. V. Best, D. Brungart, S. Carlile, C. Jin, E. Macpherson, R. Martin, K. McAnally, A. Sabin, B. Simpson: A meta-analysis of localization errors made in the anechoic free field, in Principles and applications of spatial hearing, World Scientific. 2011, pp. 14–23.
139. M.A.R. Ferreira, H. Lee: Multiscale modeling: A Bayesian perspective, Springer Series in Statistics. Springer-Verlag, New York, 2007. https://www.springer.com/gp/book/9780387708973.
140. The AMT Team: The Auditory Modeling Toolbox Full Package (version 1.x) [Code], 2021. https://sourceforge.net/projects/amtoolbox/files/AMT%201.x/amtoolbox-full-1.0.0.zip/download.

Cite this article as: McLachlan G. Majdak P. Reijniers J. & Peremans H. 2021. Towards modelling active sound localisation based on Bayesian inference in a static environment. Acta Acustica, **5**, 45.