DOI: 10.61782/fa.2023.0115

forum acusticum 2023

MODELLING DYNAMIC SOUND LOCALISATION THROUGH BAYESIAN INFERENCE: A SENSITIVITY ANALYSIS

Glen McLachlan & Herbert Peremans
University of Antwerp
Department of Engineering Management

ABSTRACT

From a Bayesian perspective, sensory information in the brain is represented in the form of probability distributions. Inherent to these probability distributions is the representation of uncertainty due to sensory noise and ambiguity. Dynamic listening using head movements is a multisensory process which involves several sources of sensory uncertainty. In this study, we introduce the numerical implementation of a Bayesian dynamic sound localisation model and investigate how the model's sensory noise parameters affect its localisation performance over the 2D sphere assuming static sound-sources in the far field. We restrict ourselves to small, open-loop head rotations. Six noise parameters that describe both acoustic and sensorimotor measurements are proposed and investigated through a sensitivity analysis. The localisation performance is expressed in lateral error, polar error and front-back confusion rate. The results from this sensitivity analysis will be compared in the future to empirical data.

Keywords: *Sound localisation, head movement, multisensory integration, Bayesian inference.*

1. INTRODUCTION

Recently, we proposed a Bayesian framework to model sound localisation that includes self motion, i.e., head movements [1]. Fundamentally, it is an extension of the static ideal-observer model presented by Reijniers et al., which used the same Bayesian theory to determine the conditional probability distribution of the sound-source direction ψ, given the available acoustic input and prior information [2].

The present implementation expands the static model on two fronts. First, the model no longer relies on a single measurement, but instead makes an observation of the available cues at set intervals during stimulus presentation [3]. This means that the the posterior distribution of the sound-source location can be recursively updated as more information becomes available. Second, the head position can be controlled over time. From this follows that not just acoustic information, but also sensorimotor information must be processed.

We previously provided a numerical implementation of the theoretical framework as a proof of concept, which qualitatively showed what information on source location could be gained from head movement [1]. However, a more in-depth analysis of its output remained to be carried out. Moreover, this proof of concept relied on two significant simplifications: 1) it encoded interaural time differences (ITDs) as the only available dynamic acoustic cue, ignoring temporal changes in spectral cues and 2) it assumed the head positions to be fully deterministic, i.e., without uncertainty.

In this paper, we remove these two simplifications and use the model to study human dynamic localisation over the full 2D sphere in the far field (i.e., direction estimation) when presented with a broadband sound-source. This model is publicly available as `"mclachlan2023"` in the Auditory Modeling Toolbox [4]. Through a sensitivity analysis we will investigate the effect of the different model parameters on localisation performance.

2. MODEL EXTENSION

2.1 Acoustic and sensorimotor information

The present model assumes the same feature space of the acoustic input static ideal-observer model by Reijniers et al. [2]: $\mathbf{y}_A$, which consists of the noiseless "true" state of the acoustic information, $\mathbf{X}_A$, convolved with noise due to uncertainty caused by the auditory system or the environment:

$$\mathbf{y}_A = [X_{itd} + \delta_{itd}, \mathbf{X}_- + \delta_-, \mathbf{X}_+ + \delta_+], \quad (1)$$

$$\mathbf{X}_- = \mathbf{X}_L - \mathbf{X}_R, \quad (2a)$$
$$\mathbf{X}_+ = [\mathbf{X}_L + \mathbf{X}_R]/2, \quad (2b)$$
$$\mathbf{X}_{L/R} = \mathbf{S} + \mathbf{H}_{L/R}, \quad (2c)$$

where $\mathbf{X}_L$ and $\mathbf{X}_R$ are the frequencywise sum log-magnitudes of the sound source spectrum, $\mathbf{S}$, and the HRTF, $\mathbf{H}_L$ and $\mathbf{H}_R$, for the left and right ear, respectively. $\mathbf{X}_-$ and $\mathbf{X}_+$ then correspond with the interaural spectral difference and an average of both monaural spectra, respectively. Note that this transformation is not strictly necessary, but aids to interpret and discuss the results. The noise sources are described as follows:

$$\delta_{itd} \sim \mathcal{N}(0, \sigma_{itd}^2) \quad (3a)$$
$$\delta_- \sim \mathcal{N}(0, \Sigma_-), \quad \Sigma_- = 2\sigma_I^2 \cdot \mathbf{I} \quad (3b)$$
$$\delta_+ \sim \mathcal{N}(0, \Sigma_+), \quad \Sigma_+ = (\sigma_I^2/2 + \sigma_S^2) \cdot \mathbf{I} + \sigma \quad (3c)$$

Variances σ_I^2, σ_S^2 and σ^2 model the noises on the spectral measurements, the subject's knowledge of the sound-source spectrum, and the cross-talk between adjacent frequency bands, respectively.

Finally, there is the sensorimotor component. Similar to $\mathbf{y}_A$, y_H denotes the noisy observation of the true state of the head orientation θ_H, which is applied to both azimuth and elevation. At each time step, θ_H is updated with a motor control signal u, which denotes a rotation of the head around the yaw or pitch axis. These variables are defined as:

$$y_H(t_i) = \theta_H(t_i) + \delta_H, \quad (4a)$$
$$\theta_H(t_{i+1}) = \theta_H(t_i) + u(t_i)\Delta t + \delta_u, \quad (4b)$$

δ_H and δ_u are the noises on the head orientation observation and the motor command, respectively. These noises ar applied to both the azimuth and elevation angles and are defined as:

$$\delta_H \sim \mathcal{N}(0, \sigma_H^2), \quad (5a)$$
$$\delta_u \sim \mathcal{N}(0, \sigma_u^2), \quad (5b)$$

Note that, for easier notation, most equations above are not described as functions of time. In reality, new measurements are made at each time step (e.g., $\mathbf{y}_A(t_i)$) and the noise variance parameters can change over time (e.g., $\sigma_{itd}(t_i)$).

2.2 Recursive Bayesian estimation

We model temporal integration of acoustic and sensorimotor information through recursive Bayesian estimation, where probability density functions (PDFs) are updated recursively over time with incoming measurements.

Following Bayes' Theorem, this process can be simply written as:

$$p_{t_i} = C \cdot p_{t_{i-1}} \cdot M, \quad (6)$$

Turning to Bayesian terminology, p_{t_i} denotes the posterior PDF, $p_{t_{i-1}}$ denotes the prior PDF and M denotes the joint sensor model which computes the likelihood. C is a normalisation constant. Note that the prior at time step t_i equals the posterior from time step t_{i-1}. At the initiation of the recursive process, $p_{t_{i-1}} = p(\psi)$, which is the spatial prior, or the prior knowledge of the sound direction. A detailed description of this equation is given in [1].

Fig. 1 demonstrates the recursive process for 5 time steps with a time step size Δt of 5 ms. Here the left two columns accumulate into an increasingly narrow distribution. We see that, despite the large variance between each look, the cumulative distribution very quickly (after 25ms) decreases in spread.

This specific example shows the process that will result in a fairly successful localisation estimate, which is presented in Fig. 2a. This does not necessarily need to happen. Fig. 2 shows the results for three different iterations of the same model parameters. If system noise causes enough incorrect observations, then localisation can be affected either by an inability to narrow the distribution (Fig. 2b), or by a narrowing of the distribution around an incorrect direction (Fig. 2c). The first effect can be considered detrimental to precision, whereas the second effect is detrimental to accuracy.

3. METHODS

A sensitivity analysis was conducted to determine how the different noise parameters affect the model's localisation performance. We considered three rotation conditions, 1527 target directions (distributed over the full sphere above an elevation of -30) and seven different individual head-related transfer functions (HRTFs), obtained in an earlier study [5]. Simulations were repeated 50 times per target direction per HRTF. For yaw and pitch rotation conditions, half of the simulations rotated the head in the positive direction and the other half rotated in the negative direction.

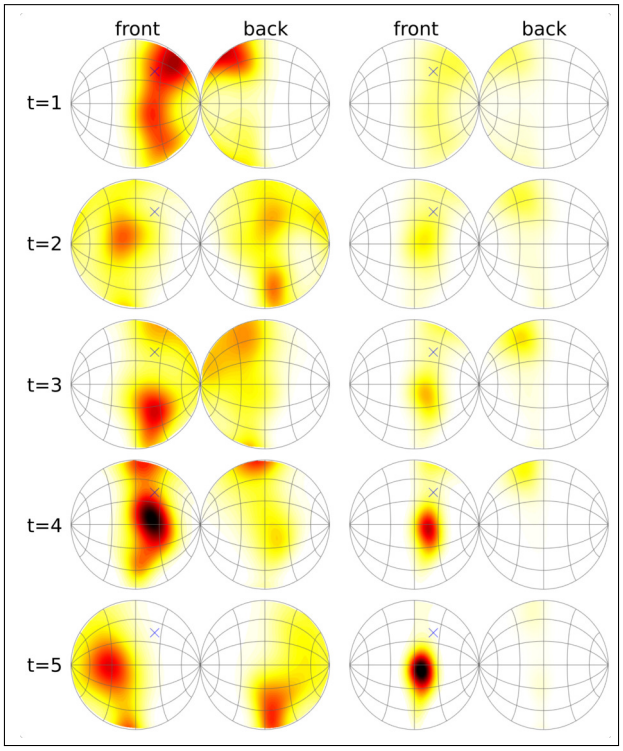

Figure 1. Probability density functions of the sound-source direction over the full sphere at time steps t=1-5, with the time between each step Δt set at 5ms. Left two columns: single look PDF at each time step. Right two columns: cumulative PDF, i.e., recursive posterior distribution at each time step. The blue 'x' marks the true source direction.

The input stimulus was a 100ms broadband white noise burst. For the movement conditions, rotations of 10° were deployed at a constant speed of $100^\circ/s$ along either the yaw or pitch axis. The initial head orientation was straight ahead, i.e., 0° azimuth and 0° elevation. The model worked with a time step size Δt of 5 ms, so the posterior was updated every 5 ms. The spatial prior was set to a uniform distribution to best visualise the effect of each tested parameter. To obtain a point estimate from the posterior PDF, we applied the maximum a-posteriori (MAP) estimate, which selects the mode of the distribution.

Localisation performance was evaluated based on three metrics: the lateral and polar root mean square error ϵ_L & ϵ_P (degrees) and the quadrant error ϵ_Q (% of trials), as defined in [6].

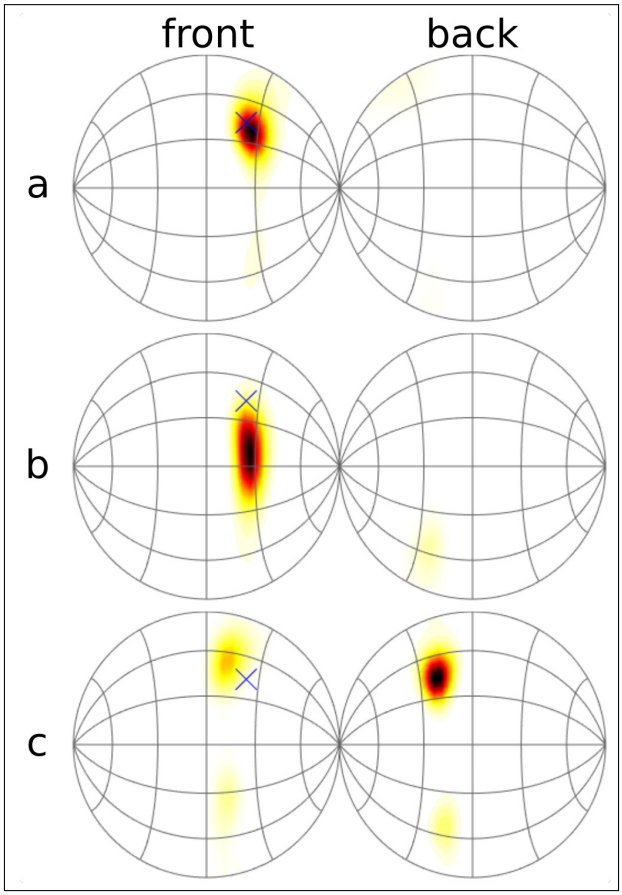

Figure 2. Three examples of probability density functions over the full sphere of the same sound-source direction at final time step t=21. a) accurate estimate, b) smeared estimate c) front-back confusion. The blue 'x' marks the true source direction.

The tested parameters and their control values are presented in Tab. 1. In preliminary simulations, it became apparent that the standard deviations of the noise models chosen in [2] were too low to provide insightful results. For this reason, the control values for acoustic noise were set large enough to see the effects of each individual parameter. Every parameter was adjusted individually, because an investigation of the interaction effects would go beyond the scope of this work. For this same reason the sensorimotor control noise was set to zero: to prevent any interaction effects between sensorimotor and acoustic uncertainty during the analysis.

Table 1. Noise parameters included in the sensitivity analysis, including descriptions of the signal they affect and their control values.

Noise on:	Symbol	Value
ITD look	σ_{itd}	3 JND
Spectrum look	σ_I	20 dB
Source knowledge	σ_S	20 dB
Head orientation look	σ_H	0°
Head motor signal	σ_u	0°

4. RESULTS

Tab. 2 presents the results of the sensitivity analysis, averaged over 7 virtual subjects, 1527 source directions and 50 repetitions. This serves as a starting point to determine the general effect of each noise parameter under different movement conditions. Fig. 3 presents the same results for ϵ_L and ϵ_P, distributed over the full sphere for the static (no movement) condition. This figure provides an insight on the direction-dependent effects of each noise source. Because the spatial effects of ϵ_P and ϵ_Q were very similar, we decided to omit figures for ϵ_Q.

Table 2. Lateral and polar root mean square error (ϵ_L, ϵ_P) and quadrant error rate (ϵ_Q) for five tested noise parameters and three rotation conditions, averaged over 7 virtual subjects, 1527 source directions and 50 repetitions. Values are rounded to one decimal place.

STATIC	ϵ_L	ϵ_P	ϵ_Q
control	4.1°	24.2°	8.1%
$2 \cdot \sigma_{itd}$	6.1°	25.4°	8.8%
$2 \cdot \sigma_I$	4.7°	30.1°	15.8%
$2 \cdot \sigma_S$	4.0°	25.9°	11.4%
$\sigma_H = 10$	4.7°	25.4°	9.3%
$\sigma_u = 2$	4.6°	21.8°	4.3%
YAW	ϵ_L	ϵ_P	ϵ_Q
control	4.6°	22.8°	5.1%
$2 \cdot \sigma_{itd}$	6.5°	24.7°	7.0%
$2 \cdot \sigma_I$	5.5°	28.8°	10.4%
$2 \cdot \sigma_S$	4.5°	24.4°	6.6%
$\sigma_H = 10$	5.3°	24.1°	6.1%
$\sigma_u = 2$	4.7°	20.7°	3.5%
PITCH	ϵ_L	ϵ_P	ϵ_Q
control	4.1°	24.3°	8.1%
$2 \cdot \sigma_{itd}$	6.1°	25.5°	8.8%
$2 \cdot \sigma_I$	4.7°	30.3°	15.9%
$2 \cdot \sigma_S$	4.1°	26.0°	11.5%
$\sigma_H = 10$	4.8°	25.4°	9.1%
$\sigma_u = 2$	4.6°	21.4°	4.0%

5. DISCUSSION

Before beginning the discussion it must be noted that the purpose of this analysis is not to match the model output to empirical data, but to test the influence of each noise source to assist parametrisation of the model in a future stage. The noise sources were divided into three categories: timing noise, spectral noise and sensorimotor noise. Each category will be discussed separately.

5.1 Control

Comparing between the three rotation conditions, it becomes apparent that yaw rotation provides a lot of information on the polar angle and the quadrant in which the sound-source is located. The decreases in ϵ_P and ϵ_Q can be explained by the hypothesis on the effect of head rotations posed in [7]. The sign of the change in ITD that accompanies a head rotation is an unambiguous indicator of the proper hemisphere, which reduces quadrant errors. Additionally, the rate of change in source azimuth angle relative to the change in head orientation can theoretically provide information on elevation angle. Pitch rotation does not affect localisation whatsoever, this agrees with the results from earlier studies [5].

5.2 Timing noise

In all rotation conditions, an increase in σ_{itd} causes a 50% increase in ϵ_L. This is no surprise: interaural cues are mostly informative about the lateral angle of a source. Perhaps more surprising are the small decreases in performance for the other two metrics. Looking at Fig. 3a, we see that these small effects mostly take place behind the listener and away from the median plane. This suggests that there are small asymmetries in the ITD between

the front and back hemispheres, which in the control condition sometimes (although rarely) provided information on the correct hemisphere. Note that the effect on ϵ_Q is larger during yaw rotation, because an increased σ_{itd} makes it more difficult to utilise the ITD rate of change, which, as mentioned before, is an indicator of the correct hemisphere.

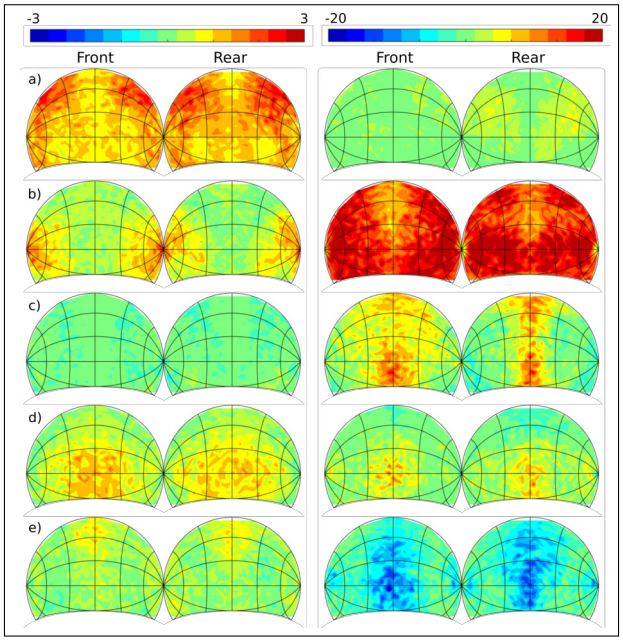

Figure 3. Model root mean square error difference between control values and separately varied model parameters, during static localisation. Left column: lateral error ϵ_L, right column: polar error ϵ_P. a) $2 \cdot \sigma_{itd}$, b) $2 \cdot \sigma_I$, c) $2 \cdot \sigma_S$, d) $\sigma_H = 10$, e) $\sigma_u = 2$. The results are plotted for 1527 target directions over the full sphere relative to the torso. Results were averaged over 7 subjects and 50 trials per subject per direction.

5.3 Spectral noise

When σ_I is increased, we see highly detrimental effects for both ϵ_P and ϵ_Q, this can be explained by the well-known role that spectral cues play in localisation along the sagittal planes. Lateral error ϵ_L is also increased, although this effect is minimised due to the complementary nature between interaural timing and level differences. For yaw rotation, an increase on spectral noise has a much smaller effect on ϵ_Q. This again demonstrates how dynamic ITD as a function of head rotation serves as a strong cue to prevent quadrant errors. The spatial analysis (Fig. 3b) shows that error increases are largest away from the median plane, this is likely because locations around the median plane are still partly supported by X_itd and $\mathbf{X}_+$, which are not or less severely affected by σ_I.

An increase in σ_S has no effect on ϵ_L, because knowledge of the source spectrum is irrelevant to lateral localisation. It also appears to be only slightly detrimental to ϵ_P and ϵ_Q. However, looking at Fig. 3c, we see that specifically directions around the median plane are much more heavily influenced than others, which contrasts to the effects in 3b. Indeed, on the median plane the listener can extract less information regarding the polar angle from $\mathbf{X}_-$ because of symmetry, and relies more on $\mathbf{X}_+$ [2]. This requires knowledge of the sound-source, which becomes uncertain as σ_S increases.

5.4 Sensorimotor noise

A higher σ_H results in an increased error for all three metrics. This is not surprising, as one would expect an uncertainty of the orientation of the head and ears to correspond with a smearing of the direction estimate. Fig. 3 shows that this smearing occurs more severely for sources close to straight ahead $(0, 0)$ and behind $(0, 180)$. This can be explained by the fact that σ_H is two-dimensional, i.e., it applies to the orientation of the head along the pitch and the yaw axis. Sources above the listener only suffer from the uncertainty in pitch and sources to the left or the right only suffer from the uncertainty in yaw. Whereas the directions at $(0, 0)$ and $(0, 180)$ are affected by both yaw and pitch uncertainty. In other words, this is an artifact of the present definition of the movement model, and empirical data would be required to determine if this adequately simulates true movements.

It may be counter-intuitive that an increase in noise on the motor control signal σ_u improves polar localisation (ϵ_P and ϵ_Q), but it can be easily explained by considering equation 4. When δ_u is high, it means that the true head orientation $\theta_H(t_{i+1})$ deviates far from the previous orientation $\theta_H(t_i)$. If this deviation is large enough, then positive effects similar to those in the yaw rotation condition can be expected. This result raises the question about the validity of localisation experiments where subjects were instructed to remain still, as it is possible that a deviation from the instructed position accidentally provided additional acoustic cues.

6. CONCLUSION & OUTLOOK

In this work we investigated the dynamic localisation model described in [1] through a sensitivity analysis of the sensory noise parameters. The current parameters already provide much control and insight on the role of acoustic and sensorimotor information during sound-localisation, but the present Bayesian framework makes it easy to implement additional elements. To conclude, we will list a number of possible adjustments, based on psychoacoustic findings. Note that this list is intended to shed a light on the potential of the recursive Bayesian framework, but is by no means exhaustive.

First, the feature space may be reformulated to be more representative of the acoustic cues that humans utilise for sound localisation. For example, several studies found that the positive spectral gradient may be a more appropriate localization cue than the absolute spectral values in each frequency band [8].

Second, lateral and polar estimation may be split into two separate processes. There is some neurological evidence that this may be the case [9], and previous work has shown that this split significantly affects the output of the Bayesian model [10]. One way of splitting this process is by applying a Bayesian decision rule depending on the plane of localization, e.g., maximum a-posteriori for the lateral angle and random sampling for the polar angle.

Third, a non-uniform spatial prior can be implemented. Empirical findings suggest that a Gaussian distribution around the horizontal plane may better describe elevation estimation [9].

Finally, earlier work showed that the introduction of a pointing error to the Bayesian estimator successfully accounted for the deviance between model and experimental data [8]. It is important to note here, however, that this pointing error cannot be included arbitrarily. To prevent that this noise is simply added to account for any deviating results, its values should be carefully chosen and grounded in empirical evidence, e.g., a higher pointing error for sources behind the listener.

7. ACKNOWLEDGEMENTS

This research was supported by the Research Foundation Flanders (FWO) under Grant number G023619N and by the Agency for Innovation and Entrepreneurship (VLAIO).

8. REFERENCES

[1] G. McLachlan, P. Majdak, J. Reijniers, and H. Peremans, "Towards modelling active sound localisation based on bayesian inference in a static environment," *Acta Acustica*, vol. 5, p. 45, 2021.

[2] J. Reijniers, D. Vanderelst, C. Jin, S. Carlile, and H. Peremans, "An ideal-observer model of human sound localization," *Biological cybernetics*, vol. 108, no. 2, pp. 169–181, 2014.

[3] P. M. Hofman and A. J. Van Opstal, "Spectro-temporal factors in two-dimensional human sound localization," *The Journal of the Acoustical Society of America*, vol. 103, no. 5, pp. 2634–2648, 1998.

[4] P. L. Søndergaard and P. Majdak, "The auditory modeling toolbox," *The technology of binaural listening*, pp. 33–56, 2013.

[5] G. McLachlan, P. Majdak, J. Reijniers, M. Mihocic, and H. Peremans, "Dynamic spectral cues do not affect human sound localization during small head movements," *Frontiers in neuroscience.-Lausanne*, vol. 17, pp. 1–10, 2023.

[6] J. C. Middlebrooks, "Virtual localization improved by scaling nonindividualized external-ear transfer functions in frequency," *The Journal of the Acoustical Society of America*, vol. 106, no. 3, pp. 1493–1510, 1999.

[7] H. Wallach, "The role of head movements and vestibular and visual cues in sound localization.," *Journal of Experimental Psychology*, vol. 27, no. 4, p. 339, 1940.

[8] R. Baumgartner, P. Majdak, and B. Laback, "Modeling sound-source localization in sagittal planes for human listeners," *The Journal of the Acoustical Society of America*, vol. 136, no. 2, pp. 791–802, 2014.

[9] R. Ege, A. Opstal, and M. M. Van Wanrooij, "Accuracy-precision trade-off in human sound localisation," *Scientific reports*, vol. 8, no. 1, pp. 1–12, 2018.

[10] H. Peremans, G. McLachlan, P. Majdak, and J. Reijniers, "Ideal versus non-ideal observer models for sound localization," in *Proceedings of the 24th International Congress on Acoustics*, 2022.