DOI: 10.61782/fa.2023.0694

forumacusticum 2023

MODEL-BASED ASSESSMENT OF SPECTRAL WEIGHTING SCHEMES FOR MEDIAN-PLANE LOCALISATION

Pedro Lladó[1*] **Piotr Majdak**[2] **Roberto Barumerli**[2] **Robert Baumgartner**[2]

[1] Department of Information and Communication Engineering, Aalto University, Finland
[2] Acoustic Research Institute, Austrian Academy of Sciences, Austria

ABSTRACT

Monaural spectral cues help humans to localise sound-sources in sagittal planes. These spectral cues mainly originate from the direction-specific filtering of the pinnae to the sound. While there is evidence that certain spectral regions contain crucial information used by the auditory system to infer the direction of the sound source, previous studies suggest that certain frequency regions are more important than others. To test this hypothesis, a sagittal-plane auditory localisation model has been modified to adopt a spectral weighting scheme. This weighting scheme assigns relative importance to the spectral cues within a frequency region to analyse its effect on predicting listener-specific patterns of localisation responses. Various model variants, which differed on their assigned spectral weights, were compared by means of Bayesian model selection. Our results show that the preferred spectral weighting was listener-specific with no clear preference at group level. Thus, our findings suggest that listeners apply different decoding strategies of pinna cues for sound localisation.

Keywords: *Localisation, spectral cues, auditory model*

1. INTRODUCTION

Human listeners use monaural spectral cues to localise sound sources in sagittal planes, i.e. up-down and front-back. These cues originate from the direction-dependent acoustic filtering of the torso, the head and mainly the pinnae [1–3]. The direction-dependent filtering affects the spectrum of the sound, being perceptually important for sagittal-plane localisation in the range 0.7 to 18 kHz [4,5].

The exact mechanism of how spectral cues are integrated over frequency is not fully understood yet. For example, in order to address spectral cues integration over frequency, Zonooz et al. studied the contribution of the most prominent spectral notch in the HRTF on elevation, since its center frequency tends to present a monotonic behavior over polar angle [6]. Their results show that the main notch region contribution is important for elevation perception, but they also found that neighbouring regions play an important role. In another study, Ahrens and Brimijoin conducted a behavioral task in which the participants were asked whether a target sound was above or below a given reference [7]. The target sound was composed of multiple narrow-band noises, each of them convolved with a randomised elevation of their HRTF and presented at the same time. From the collected data, they analysed which frequency bands contributed more to the actual response of the subjects. The results of these two studies resulted in different contributions of each spectral region.

We analyse here the frequency-dependent contribution of spectral cues for sagittal plane localisation using an auditory model [5]. To that end, three model variants were defined, which differed in their spectral weighting scheme, and they were tested on their ability to predict individual localisation data by estimating the parameters with a maximum-likelihood optimisation procedure. We implemented a model comparison approach based on Bayesian statistics to analyse the results of each model variant on their ability to predict actual responses both at

**Corresponding author*: *pedro.llado@aalto.fi.*

an individual and group level.

2. SPECTRAL WEIGHTING SCHEMES

The model proposed by Baumgartner et al. [5] compares the spectral cues of the target sound to the listener's template. The templates are pre-computed for each direction of a given sagittal plane as a set of spectral cues derived from the directional transfer functions (DTF) for each auditory band. The same computation is applied to the target sound. Thus, a distance function δ averages over auditory bands the difference between the target spectral representation and each available direction in the template. From these distances, a probability mass vector (PMV) is computed that represents the probability of the listener responding to each polar angle given a target sound. To enable the comparison of spectral weighting schemes, we modified the distance function δ to perform a weighted average instead.

Three different sets of normalised weights were considered as three model variants. The first model variant applied equal weighting across the entire frequency range, as in the original model from Baumgartner et al. [5]. The second one follows the spectral weighting scheme proposed by Zonooz et al. [6], which assigns large weights to the most-prominent spectral notch region (6 - 9 kHz) and the weights are non-zero only between 2.5 and 12 kHz. The third model variant used the results from the study conducted by Ahrens and Brimijoin [7], which assigned non-zero weights between 1 and 15 kHz (except for the 1.6 kHz band) and presents a peak at 6.5 kHz. We will refer to each model variant as Ω_B, Ω_Z and Ω_A, respectively, from their first-authors surname. The weights for each model variant are shown in Figure 1.

We assessed the three model variants on their ability to predict the localisation responses of 18 subjects from the baseline condition in Baumgartner et al. [5]. Their individually measured DTFs (from the ARI HRTF-Database) and their localisation responses were used, which are available in the Auditory Modelling Toolbox [8]. The measurement setup was the same for all the studied listeners and is described in [9]. The quality of these measured DTFs was considered sufficient based on the listeners' performance in a localisation task. These measured DTFs were used both as the template and the target in the model evaluation, as in the baseline condition in Baumgartner et al. [5].

To compare the model variants, we first performed a parameter estimation procedure based on maximum-likelihood optimisation. Then, we compared the model variants by means of the Bayesian Information Criterion (BIC) on their ability to predict the localisation responses. These BICs were computed for each model variant and listener, and then compared using a Bayesian model comparison technique. This Bayesian model comparison was performed using the Statistical Parametric Mapping toolbox [10, 11]. Importantly, this model comparison approach outputs the probability of selecting one model compared to the rest for each listener. From the listener-dependent probability, a group level probability of selecting each model is computed, together with the Bayesian Omnibus Risk (BOR), i.e. probability that the compared models are equally good at explaining the data. The selection probability corrected by the BOR results in the protected exceedance probability (PXP), which accounts for the possibility of the results being due to chance [10].

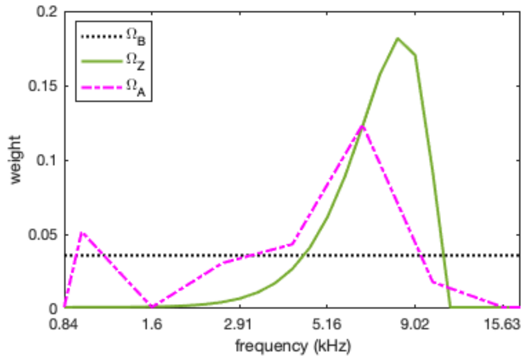
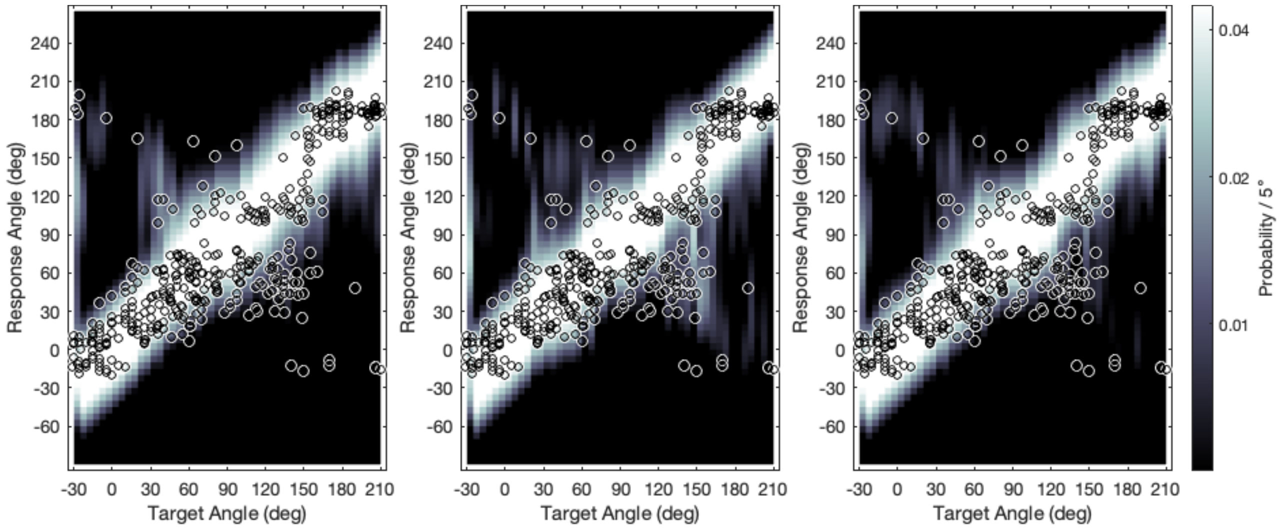


Figure 1: Spectral weights applied by each model variant.

3. RESULTS AND DISCUSSION

The spectral weighting schemes had an effect on the PMVs. In Figure 2, the PMVs obtained for an exemplary subject (NH16) are compared to their actual responses collected in a listening experiment. This example aims at representing the differences introduced to a PMV by each of the model variants. Upon visual inspection, the model variant Ω_Z predicted a higher probability of back-to-front confusions for target angles between 90° and 180°. This pattern was also found in the actual responses, resulting in a better prediction.

At group level, the PXPs for each model were: PXP(Ω_B) = 0.37, PXP(Ω_Z) = 0.32 and PXP(Ω_A) = 0.31;



(a) Ω_B (b) Ω_Z (c) Ω_A

Figure 2: PMVs for an exemplary listener (NH16) predicted by each model variant. Probabilities are encoded by brightness according to the color bar to the right. Actual responses are shown as open circles.

and the BOR = 0.86. These results suggest that none of the model variants was better at explaining the data at group level. This could be explained by the listener-dependent model selection. The model variant Ω_B was selected as the best for 8/18 listeners (with a model variant probability p>0.75). The model variant Ω_Z was best for 5/18 listeners (p>0.75) and the model variant Ω_A was best for 5/18 listeners (p>0.75) (see Figure 3).

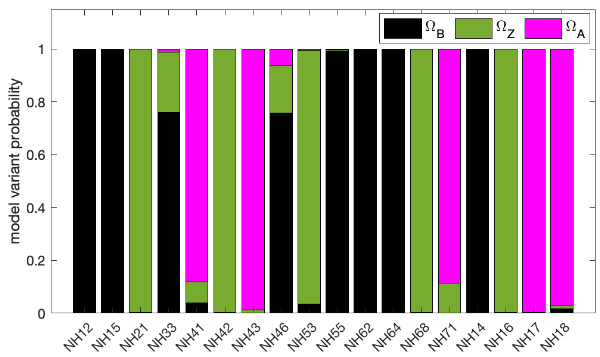


Figure 3: Listener-specific probabilities for model variant selection.

Our results show that there was a high risk (BOR = 0.86) on selecting a model at group level. Therefore, substituting the original non-weighted average δ function by a weighted one based on Ω_Z or Ω_A is not justified given our data. On the one hand, an interpretation of the high BOR is that we did not include in the comparison a weighting scheme that represents the actual spectral integration mechanism. Such weighting scheme would, therefore, improve the model predictions at group level.

Another interpretation is that the spectral weights are listener dependent per se. A plausible explanation relies on the assumption that each listener learns to integrate the spectral cues, e.g. depending on the spectral information that they have access to. If a listener's pinnae provide unambiguous cues to differentiate elevation at a certain frequency range, it is reasonable to think that they would learn to exploit these cues. On the other hand, if a frequency region provides confusing localisation cues, they would learn to discard them. However, this hypothesis needs to be studied in depth before any conclusions are made.

4. CONCLUSION

We modified a sagittal-plane auditory localisation model to adopt a spectral weighting scheme, which assigns relative importance to the spectral cues within a frequency region. Three model variants, which differed on their spectral weighting scheme, were compared on their ability to predict localisation data. The results showed that modifying the spectral weighting scheme had an effect on the model predictions. The Bayesian model selection results suggested that none of the model variants was better than the rest at explaining the data at group level. However, each model variant was best at predicting the localisation data for a subset of listeners. Even though a deeper analysis is needed, the results suggest that different listeners may apply a different decoding strategy for pinna cues decoding.

5. REFERENCES

[1] J. Hebrank and D. Wright, "Spectral cues used in the localization of sound sources on the median plane," *The Journal of the Acoustical Society of America*, vol. 56, no. 6, pp. 1829–1834, 1974.

[2] J. C. Middlebrooks, "Narrow-band sound localization related to external ear acoustics," *The Journal of the Acoustical Society of America*, vol. 92, no. 5, pp. 2607–2624, 1992.

[3] J. Blauert, *Spatial Hearing: the Psychophysics of Human Sound Localization*. MIT press, 1997.

[4] V. R. Algazi, C. Avendano, and R. O. Duda, "Elevation localization and head-related transfer function analysis at low frequencies," *The Journal of the Acoustical Society of America*, vol. 109, no. 3, pp. 1110–1122, 2001.

[5] R. Baumgartner, P. Majdak, and B. Laback, "Modeling sound-source localization in sagittal planes for human listeners," *J. Acoust. Soc. Am.*, vol. 136, pp. 791–802, Aug. 2014.

[6] B. Zonooz, E. Arani, K. P. Körding, P. Aalbers, T. Celikel, and A. J. Van Opstal, "Spectral weighting underlies perceived sound elevation," *Scientific reports*, vol. 9, no. 1, pp. 1–12, 2019.

[7] A. Ahrens and O. Brimijoin, "Perceptual weighting of elevation localization cues across frequency," in *Poster session presented at 177th Meeting of the Acoustical Society of America, Louisville, Kentucky, United States*, 2019.

[8] P. Majdak, C. Hollomey, and R. Baumgartner, "Amt 1. x: A toolbox for reproducible research in auditory modeling," *Acta Acustica*, vol. 6, p. 19, 2022.

[9] P. Majdak, M. J. Goupell, and B. Laback, "3-d localization of virtual sound sources: Effects of visual environment, pointing method, and training," *Attention, perception, & psychophysics*, vol. 72, no. 2, pp. 454–469, 2010.

[10] L. Rigoux, K. Stephan, K. Friston, and J. Daunizeau, "Bayesian model selection for group studies — revisited," *NeuroImage*, vol. 84, pp. 971–985, 2014.

[11] K. E. Stephan, W. D. Penny, J. Daunizeau, R. J. Moran, and K. J. Friston, "Bayesian model selection for group studies," *NeuroImage*, vol. 46, no. 4, pp. 1004–1017, 2009.