DOI: 10.61782/fa.2023.0696

forum acusticum 2023

MODELLING NARROW-BAND SOUND LOCALISATION IN THE MEDIAN PLANE USING SPECTRAL GRADIENTS

Pedro Lladó[1*] **Petteri Hyvärinen**[1] **Ville Pulkki**[1]

[1] Department of Information and Communications Engineering, Aalto University, Finland

ABSTRACT

Localisation of narrow-band sounds is heavily influenced by their frequency content. Middlebrooks [1] demonstrated the relation between the perceived location of a sound and the subjects' directional transfer function. Since then, neurophysiological studies have shown evidence of the sensitivity of the dorsal cochlear nucleus to the positive gradient of the sound spectrum, offering a physiologically-inspired alternative for modelling auditory localisation. In an attempt at connecting narrow-band localisation data to the modern broad-band localisation auditory models, the current work presents recent perceptual results, which have been analysed using the positive gradient of the spectral profile. Results obtained by the spectral gradient analysis were similar to the ones obtained using the directional transfer function. Thus, using the positive gradient of the spectrum as a proxy for the internal representation of the spectral cues allows the modelling of both narrow-band and broad-band auditory localisation data with the same computational approach.

Keywords: *Sound localisation, narrow-band, auditory model.*

1. INTRODUCTION

Sound source localisation helps humans analyse their surroundings. Localisation of sound sources in the lateral dimension, left-right, is in general robust due to interaural cues, i.e. interaural time and level differences [2–4]. On the other hand, localisation in the polar dimension, i.e. front-back and up-down, is less accurate as it relies on monaural spectral cues [4–7] and on the dynamic integration of interaural cues [4, 8, 9]. When the target sound is short so that we cannot use the dynamic information, the perceived direction is derived solely from monaural cues. These monaural cues emerge from to the direction-dependent effect of the torso, the head and mainly the pinnae to the spectrum of the arriving sound.

Localisation of narrow-band sounds presents a peculiarity that requires special attention. Humans tend to localise narrow-band sounds in the polar dimension based on their frequency content instead of their actual location [1, 4, 10]. If the spectrum of the sound is limited within a single auditory filter, moving the sound along the polar dimension does not result in perceivable spectral differences. Thus, it seems reasonable to assume that the bandwidth of a sound needs to be broader than an auditory filter to resolve the direction of arrival from its spectrum. This peculiarity has motivated several experiments to understand narrow-band sound localisation [1, 7, 10–13].

One of these studies discovered a key aspect in our current understanding of narrow-band sound localisation [1]. In this study, Middlebrooks found a connection between the center frequency of a narrow-band sound and its perceived direction in the polar domain. They demonstrated that listeners localised a narrow-band stimulus at the angle where their directional transfer function (DTF) presented a maxima at the stimulus center frequency. A model based on this principle was able to predict narrow-band localisation data [1], and therefore this relation is widely accepted.

However, it is possible that mammals do not have ac-

*__Corresponding author__: pedro.llado@aalto.fi.

cess to information about the absolute magnitude of the spectral components to resolve the sound source direction. Evidence found after [1] suggests that mammals are sensitive to the gradient of the spectrum instead of its magnitude. More precisely, Reiss and Young [14] showed that the dorsal cochlear nucleus (DCN) of cats was sensitive to the positive spectral gradient of the presented stimuli. Even though this has only been demonstrated in small mammals, a model based on the positive spectral gradient (PSG) was able to predict human localisation data of broad-band sounds for various listening conditions [6, 15, 16]. If this assumption is true and we do not have access to information about the magnitude of the spectrum for broad-band sounds, modeling polar localisation behavior should rely on PSG information only.

In this study, we analyse if access to the magnitude of the spectrum is needed to predict the behavioral responses to narrow-band sounds. If this prediction is possible from the PSG, we could explain both broad-band and narrow-band auditory localisation data with the same computational approach. We used an auditory model [6] to predict actual localisation responses using two model variants. The first one used a classical approach and had access to the magnitude spectral profile of the DTF. The second variant had access to the PSG only. We compared the ability of both model variants on estimating median plane localisation data to determine whether two different approaches are needed depending on its bandwidth.

In section 2, the data from Kim et al. [10] is introduced, which will serve as the basis for the model variants comparison. The model variants are compared in Section 3, both at a group level and for individual subjects. Section 4 discusses how the results may fit within the previous theoretical framework.

2. PERCEPTUAL DATA

The perceptual data analysed in this manuscript was collected by Kim et al. [10]. They conducted a narrow-band localisation test for sound sources located in the median plane. Nine subjects participated in the localisation of 1/3 octave band-passed noise. Each stimulus consisted of two noise bursts of 200 ms length, separated by 800 ms, where both noise bursts were of the same center frequency and source location. The target sounds were presented from the front, top and back loudspeakers (θ_t = 0°, 90° and 180°). The center frequencies varied between 125 Hz and 12.5 kHz and the sound pressure level was randomised between 69 and 71 dB SPL, A-weighted. The subjects were asked not to move their heads and they responded the perceived polar angle of the target sound using a circular slider on a touchscreen. Each subject responded to eight sound stimuli from each direction and center frequency, for a total of 528 stimuli in a randomised order.

The results showed a spectral region between 2.5 kHz and 8 kHz, where the perceived location depended on the center frequency of the target sound. Within this spectral region, variance of the responses decreased and the median values were consistent across different source directions and subjects. The responses in this frequency range showed a monotonic increasing curve from about 20° to 90° for increasing values of center frequency.

3. AUDITORY MODEL ANALYSIS

An auditory model for sagittal plane localisation, proposed by Baumgartner et al. [6], was used to predict the behavioral data from [10]. We examined with this auditory model, available in [17], whether the PSG approach could be an alternative to the magnitude response peaks on explaining narrow-band localisation data.

First, the model approximates the processing of the cochlea by filtering the sound using a gammatone filterbank [18]. Then, the spectral magnitude profile ξ, in dB, is computed by averaging over time the output of each gammatone filter. The PSG γ is computed as:

$$\gamma[b] = \max(\xi[b] - \xi[b-1], 0), \tag{1}$$

where b is the frequency band.

The template of a listener contains the spectral representation γ for each available direction in their measured DTF. The target sound is obtained by convolving the stimulus, i.e. the narrow-band noise, with the target direction of the DTF. The spectral representation γ of the target sound is computed following the same procedure as for one direction in the DTF. For each direction in the DTF, a distance metric δ is computed between the spectral representation of the target sound and the template. From these distances, the similarity to each direction in the template is computed with a sigmoid psychometric function. The similarities are then smeared using a Gaussian function to simulate the mapping between the auditory perception and the motor response. Then, the smeared similarities are normalised into a probability mass vector (PMV). The PMV represents the probability of a listener responding to each polar angle given the target sound.

To test our hypothesis, two model variants were defined. The first one uses γ as the spectral representation of

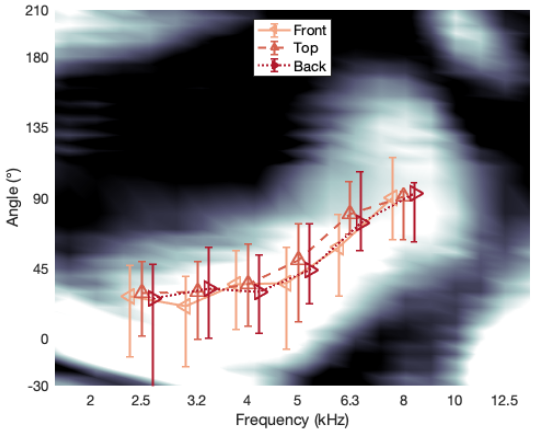
sound, which bases on the evidence from [14] and follows the description above, as proposed by [6]. We will refer to this model variant as Ω_γ. The second model variant is adopted from [1] and uses the spectral magnitude profile ξ as the spectral representation of sound ($\gamma = \xi$). Thus, the PSG is not computed and the distance metric δ is applied directly to the spectral magnitude profie ξ. We will refer to this model variant as Ω_ξ.

3.1 Modelling spectral cues for a pool of subjects

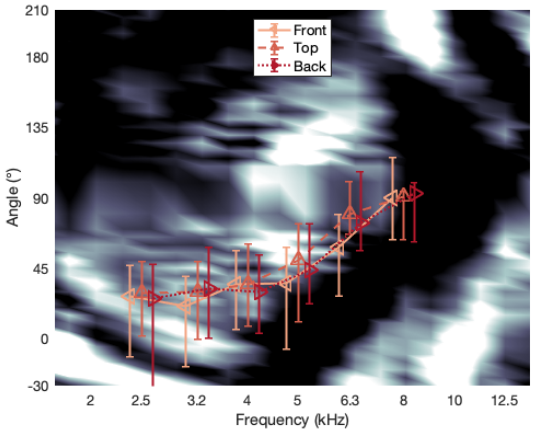
Middlebrooks found that the peak of the magnitude spectral profile is connected to the perceived location of narrow-band sounds [1]. Thus, the peaks in the spectral cues representation from the model variant Ω_ξ should correspond with the localisation responses from a behavioral task. Ideally, this correspondence should be examined statistically, comparing the individual behavioral results with the output of the individualized model for each subject. Due to the low number of participants in [10] for whom both behavioral data and DTFs were available, behavioral and model data were averaged over individuals and the correspondence was inspected visually. The spectral cues representation for the pool of 23 subjects used in [6] were computed.

The results of the average spectral magnitude profile and the group-level behavioral data are shown in Figure 1a. The illustration shows that the spectral region between 2.5 kHz and 8 kHz follows the shape of the DTF maxima at a group level, as found in the behavioral task [10]. Thus, the qualitative comparison between the predicted and behavioral data for group-level responses is in accordance with [1]. Our analysis with new data agrees on the possibility of explaining narrow-band localisation from the spectral magnitude profile of the DTFs.

The same comparison procedure was performed for the model variant Ω_γ. This time, the localisation data obtained by [10] was compared qualitatively to the spectral cues representation after extracting the PSG. The results are shown in Figure 1b. From visual inspection, the spectral cues representation seem to also follow the trend of the responses from the behavioral task. In this case, the localisation data seems to fall around the lowest polar angle with non-zero PSG at the studied frequencies. Concluding if this mechanism is present in human narrow-band localisation is out of the scope of this manuscript, but it seems plausible that the model variant Ω_γ could contain enough information to provide similar results as Ω_ξ.



(a) Without PSG



(b) With PSG

Figure 1: Black and white canvas: probability mass vectors for both model variants at a group level (average). The spectral representation is encoded by brightness: in a), brighter represents higher magnitude in ξ; in b), brighter represents higher PSG in γ. Orange distributions: interquartile range for the perceptual data responses at a group level (each tone of orange represents a different θ_t)

.

3.2 Modelling responses for individual subjects

Individual DTFs were available for five of the subjects in the behavioral experiment from [10]. We used the model from Baumgartner et al. [6] to estimate their responses for both model variants, without the PSG extraction, Ω_ξ, and with the PSG extraction, Ω_γ.

The input stimuli for the model were 1/3 octave band-passed filtered noise, as in the behavioral experiment [10]. The region from 2.5 kHz to 8 kHz delimited the frequency range for which the stimuli were generated, based on the perceptual results. These stimuli were convolved with the DTF of the subjects for the front, top, and back direction, to mimic the procedure in [10]. The model parameters degree of selectivity (Γ = 6 dB^{-1}) and response scatter (ε = 17°) were already optimised at a group level in the original model [6] and the same values were used here. The listener-specific sensitivity could not be optimised for each subject here due to the lack of responses besides the three studied source locations. Therefore, the default value was used (S = 1). This simplification is thought to enable a valid comparison, since we did not aim at predicting absolute metrics for each subject. We assume here that a modification of the S value would affect both model variants similarly.

To avoid noise interfering in the template to target comparison, we implemented a signal detection stage that consisted on a threshold defined as -30 dB from the maximum in the spectral magnitude profile. The bands with a magnitude value below this threshold were discarded for the distance metric δ calculation. Since the magnitude of δ was different between model variants, it was normalised to guarantee a fair comparison. The values of δ were scaled to range between 0 and 10 to maintain similar values to those range to those obtained in the model variant Ω_ξ with non-normalised δ .

The PMVs for both model variants and the response data from the behavioral experiment are shown in Figure 2. For the purpose of comparing each model's ability to explain the behavioral data, the log-likelihood was computed:

$$\mathcal{L}_{\Omega v} = \sum_{n=1}^{\mathrm{N}} \ln\left[\mathrm{p}_{\Omega_v}(\theta_{\mathrm{r}_n})\right], \tag{2}$$

where N is the number of responses, θ_{r} is the polar angle of the actual response, v is the model variant, i.e. γ or ξ, and p is the probability from the PMV. The Bayesian Information Criterion (BIC) was computed to summarise the model variant's performance:

$$\mathrm{BIC}_{\Omega_v} = \mathrm{k} \cdot \ln \mathrm{N} - 2 \cdot \mathcal{L}_{\Omega_v}, \tag{3}$$

where k is the number of fitted parameters in the model (k = 3, fitted at a group level). The BIC for each subject and model variant is presented in Table 1.

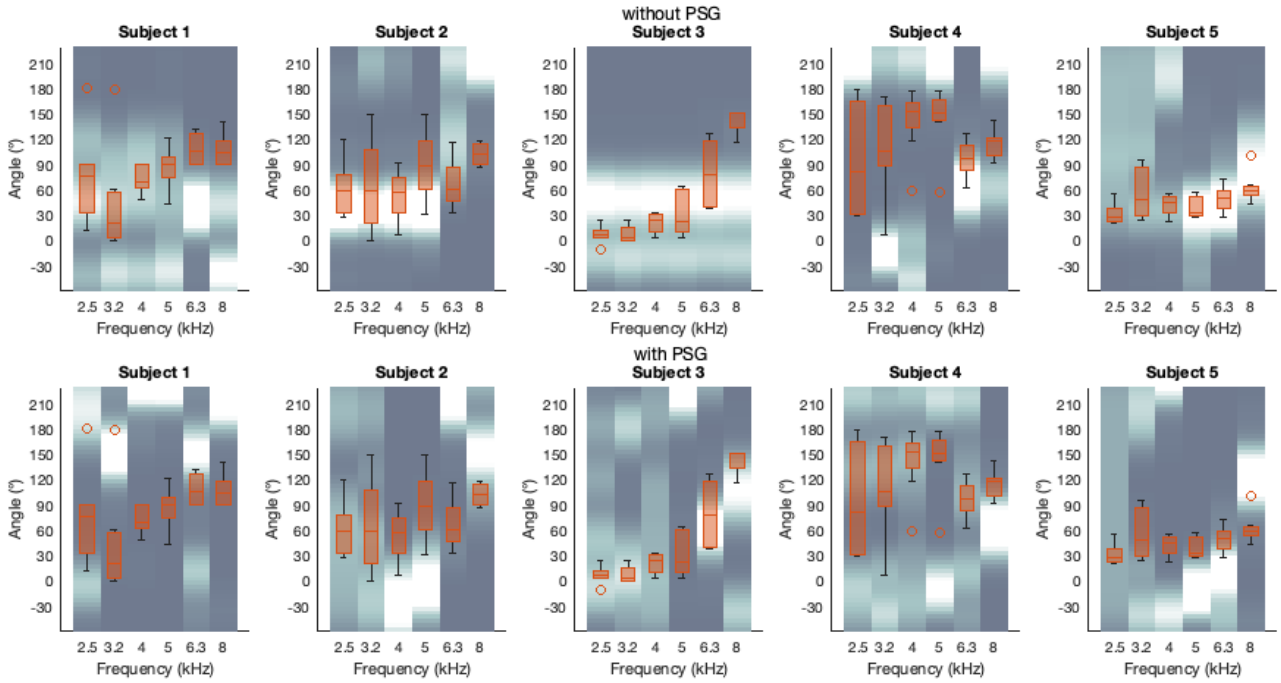
The performance of each model variant was inconsistent across listeners (see Table 1). Subject 5 is an exemplary listener for whom the model variant Ω_ξ predicts better their responses than the model variant Ω_γ (see Figure 2). In contrast, Subject 2 is an exemplary listener for whom the model variant Ω_ξ predicted probabilities do not match with their actual responses. However, the model variant Ω_γ predictions do not seem to follow the actual response patterns either.

The Bayesian Omnibus Risk (BOR), defined as the probability that the model variants are equally competent in describing the data, was BOR = 0.61 on average after ten iterations (std $<$ 0.01) [19, 20]. While a BOR value that confirms statistical significance is not defined by method described in [20], we consider the obtained BOR to be high. Thus, with our data we can't determine which model variant is better.

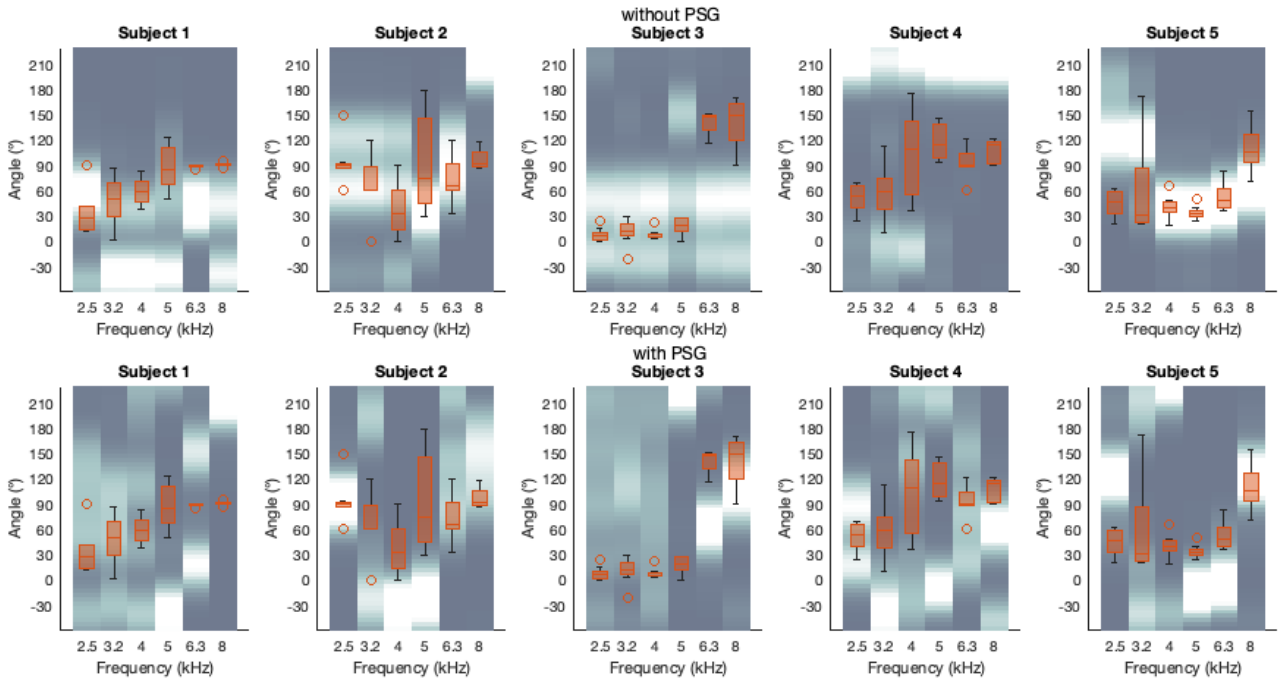
4. DISCUSSION

This study investigated whether narrow-band localisation can be modeled by means of the PSG instead of the spectral magnitude profile. In general terms, the model predictions did not show a good match with the individual perceptual data. To guarantee better model predictions, future work should perform a listener-dependent optimisation of the model parameter S. For this, a larger amount of responses would be required, which would eventually improve the quality of the perceptual data. Furthermore, based on a visual inspection of the Figure 2, some subjects seem to suffer from an elevation bias. If this could be confirmed and corrected, both model variants could eventually predict better the actual responses from the behavioral task. These possible improvements should be taken into account for a more exhaustive analysis of the data and for a more fair and robust comparison. Even though the model predictions were modest, both model variants presented similar limitations, and the comparison was therefore considered fair.

The perceptual data was collected in a previous localisation test by [10]. They found a spectral region that falls within a monotonically increasing region of the magnitude spectral profile of the DTFs at a group level (see

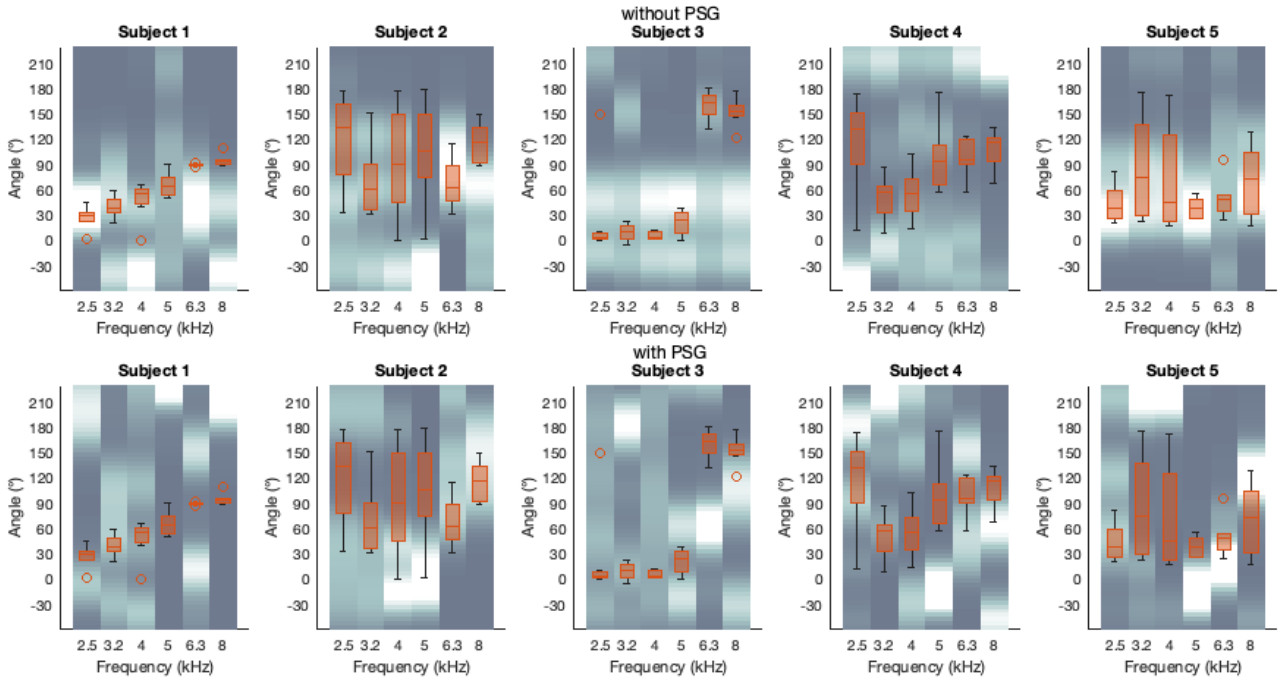


(a) Model PMVs and behavioral responses for $\theta_t = 0°$.



(b) Model PMVs and behavioral responses for $\theta_t = 90°$.

Figure 2: Black and white canvas: Listener-specific PMVs at the model output for each virtual 1/3 octave band noise stimulus. The probabilities are encoded by brightness, where brighter represents higher probability. Orange boxplots: distributions of actual responses for each subject in the behavioral experiment. (Figure continues in the next page.)



(c) Model PMVs and behavioral responses for $\theta_t = 180°$.

Figure 2: Black and white canvas: Listener-specific PMVs at the model output for each virtual 1/3 octave band noise stimulus. The probabilities are encoded by brightness, where brighter represents higher probability. Orange boxplots: distributions of actual responses for each subject in the behavioral experiment.

Figure 1a). This is in accordance to [1] regarding the link between the DTF and the perceived location of narrow-band sounds. This seems to provide a good starting point for the comparison performed in this study.

To compare whether the PSG contained sufficient information for narrow-band localisation, an auditory model for sagittal plane localisation was used [6]. The connection between the two studied model variants is defined in Equation 1. The PSG can be understood as a local extrema detector, i.e. peak and notch finder. Therefore, it was expected that the PSG is able to track the maximum of the magnitude spectral profile. Nonetheless, the PSG does not have access to information about absolute extrema. This means that the maximum of the spectral magnitude profile could be hidden among noise, and therefore uninformative regarding the connection to the perceived location. Figure 1b shows a trend where the responses of the subjects in [10] correspond to the lowest polar angle where the PSG is positive for the pool of subjects. Even though we hypothesised that the PSG might contain enough information to explain narrow-band localisation, the connection the lowest polar angle was not anticipated. Future work should aim at analysing if this connection is maintained at a subject level with a larger number of participants.

The subject-dependent analysis based on Bayesian statistics resulted in a high BOR. In other words, both model variants are similarly competent at explaining our experimental data. While this model variants comparison should be tested more exhaustively in the future, our results suggest that the modelling approach should be consistent independently of the bandwidth of the sound stimulus.

Our results suggest that there is potential in explaining narrow-band localisation by PSG analysis. Based on our results, it seems that Middlebrooks' theory holds even after the new evidences on the PSG found by Reiss and Young [14]. Therefore, it seems reasonable to use the same modelling approach independently of the bandwidth

Table 1: Bayesian information criterion (BIC) on predicting the actual responses for each target direction θ_t (average over 10 iterations). Lower BIC values reflect more accurate predictions of the responses.

Subject	θ_t	BIC_{Ω_ξ}	BIC_{Ω_γ}
1	0°	403.9	533.0
	90°	382.8	432.2
	180°	370.0	430.0
	all	1158.3	1225.7
2	0°	421.0	417.4
	90°	388.4	482.8
	180°	401.8	466.1
	all	1125.9	1146.9
3	0°	424.6	494.3
	90°	470.0	451.6
	180°	493.1	440.1
	all	1286.7	1260.7
4	0°	648.0	432.6
	90°	637.3	431.8
	180°	630.7	428.1
	all	1944.7	1200.0
5	0°	404.2	439.5
	90°	362.8	413.1
	180°	367.7	413.1
	all	1008.6	1196.7

of the sound. However, no final conclusions should be made before conducting further experiments.

5. REFERENCES

[1] 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.

[2] L. Rayleigh, "Xii. on our perception of sound direction," *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, vol. 13, no. 74, pp. 214–232, 1907.

[3] E. A. Macpherson and 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*, vol. 111, no. 5, pp. 2219–2236, 2002.

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

[5] 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.

[6] 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.

[7] 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.

[8] S. Perrett and W. Noble, "The contribution of head motion cues to localization of low-pass noise," *Perception & psychophysics*, vol. 59, no. 7, pp. 1018–1026, 1997.

[9] H. Pöntynen and N. H. Salminen, "Resolving front-back ambiguity with head rotation: The role of level dynamics," *Hearing research*, vol. 377, pp. 196–207, 2019.

[10] T. Kim, H. Pöntynen, V. Pulkki, *et al.*, "Directional-band-dominant spectral region for the sound localisation in the median plane," in *International Congress on Acoustics*, 2022.

[11] J. Blauert, "Sound localization in the median plane," *Acta Acustica united with Acustica*, vol. 22, no. 4, pp. 205–213, 1969.

[12] M. Itoh, K. Iida, and M. Morimoto, "Individual differences in directional bands in median plane localization," *Applied Acoustics*, vol. 68, no. 8, pp. 909–915, 2007.

[13] R. Wallis and H. Lee, "Directional bands revisited," in *Audio Engineering Society Convention 138*, Audio Engineering Society, 2015.

[14] L. A. Reiss and E. D. Young, "Spectral edge sensitivity in neural circuits of the dorsal cochlear nucleus," *Journal of Neuroscience*, vol. 25, no. 14, pp. 3680–3691, 2005.

[15] R. Baumgartner and P. Majdak, "Modeling localization of amplitude-panned virtual sources in sagittal planes," *J. Audio Eng. Soc.*, vol. 63, p. 562, Aug. 2015.

[16] R. Baumgartner, P. Majdak, and B. Laback, "Modeling the effects of sensorineural hearing loss on sound localization in the median plane," *Trends in Hearing*, vol. 20, pp. 1–11, Sep. 2016.

[17] P. Majdak, C. Hollomey, and R. Baumgartner, "Amt 1.0: The toolbox for reproducible research in auditory modeling," *submitted to Acta Acustica*, 2021.

[18] R. F. Lyon, "All-pole models of auditory filtering," *Diversity in auditory mechanics*, pp. 205–211, 1997.

[19] 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.

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