
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Lladó, Pedro; Hyvärinen, Petteri; Pulkki, Ville

Auditory model-based estimation of the effect of head-worn devices on frontal horizontal localisation

Published in:
Acta Acustica

DOI:
[10.1051/aacus/2021056](https://doi.org/10.1051/aacus/2021056)

Published: 06/01/2022

Document Version
Publisher's PDF, also known as Version of record




Published under the following license:
CC BY

Please cite the original version:
Lladó, P., Hyvärinen, P., & Pulkki, V. (2022). Auditory model-based estimation of the effect of head-worn devices on frontal horizontal localisation. *Acta Acustica*, 6, 1-13. Article 2021056. <https://doi.org/10.1051/aacus/2021056>

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.



Auditory model-based estimation of the effect of head-worn devices on frontal horizontal localisation

Pedro Lladó^{*} , Petteri Hyvärinen , and Ville Pulkki 

Department of Signal Processing and Acoustics, Aalto University School of Electrical Engineering, Otakaari 5A, 02150 Espoo, Finland

Received 29 December 2020, Accepted 9 December 2021

Abstract – Auditory localisation accuracy may be degraded when a head-worn device (HWD), such as a helmet or hearing protector, is used. A computational method is proposed in this study for estimating how horizontal plane localisation is impaired by a HWD through distortions of interaural cues. Head-related impulse responses (HRIRs) of different HWDs were measured with a KEMAR and a binaural auditory model was used to compute interaural cues from HRIR-convolved noise bursts. A shallow neural network (NN) was trained with data from a subjective listening experiment, where horizontal plane localisation was assessed while wearing different HWDs. Interaural cues were used as features to estimate perceived direction and position uncertainty (standard deviation) of a sound source in the horizontal plane with the NN. The NN predicted the position uncertainty of localisation among subjects for a given HWD with an average estimation error of 1°. The obtained results suggest that it is possible to predict the degradation of localisation ability for specific HWDs in the frontal horizontal plane using the method.

Keywords: Auditory model, Horizontal localisation, Subjective evaluation, Head worn device, Neural network

1 Introduction

Head-worn devices (HWDs), such as helmets or hearing protectors, may obstruct the acoustic transmission of sound to the ear completely or partially. In potentially hazardous environments, workers often need to wear protection HWDs and it is vital to ensure that the performance or safety of the person wearing the device is not unnecessarily compromised due to any potential deterioration of auditory perception. Depending on the task of the worker, the ability to communicate verbally, locate and identify potential objects of interest, and maintain a sense of space may be essential for successfully carrying out their operations. A traditional perceptual evaluation of a HWD requires conducting structured listening tests which can require a large number of participants and substantial time to prepare and complete. Especially during the development of new products, large scale listening tests for each potential feature are not a viable option.

Acoustic measurements, on the other hand, are relatively easy and quick to perform. In principle, the head-related transfer functions (HRTFs) from a source to the ear canals of a subject can be measured with and without a HWD, and the differences between the responses can then be monitored. However, it is not a straightforward task to

evaluate the effect of a device on the spatial perception of surroundings from these differences. Developing tools that provide robust evaluation from acoustical measurements would contribute to improving product quality and reducing development time and costs. If the evaluation estimates the subjective effect introduced by the product accurately, intermediate modifications in early stages of development can be tested without the need of subjective experiments.

In this study, an instrumental method based on acoustic measurements is proposed to assess how HWD products introduce distortions to the binaural cues estimates. This method aims at contributing to a larger framework to assess the sound quality of HWDs, including front-back confusions, elevated sources, externalisation, timbral differences, etc. The degradation caused by HWDs on horizontal localisation was assessed with formal subjective listening experiments, and a method to estimate the degradation with a computational model-based analysis of acoustically measured responses was developed. The method estimated the binaural cues with a simple binaural auditory model and utilised a neural network (NN) to estimate the spatial perception with the cues as input features. The scope of this model was to provide a method for the objective assessment of a HWD performance, based on subjective data. This model may be useful for applications such as characterizing HWDs, evaluating hear-through systems in virtual reality, or improving the methods for evaluating the acoustic

^{*}Corresponding author: pedro.llado@aalto.fi

transparency of audio devices. The method is thus an application of a computational model of binaural hearing to a real-world problem, similar to several methods in [1].

2 Background

2.1 Models of binaural hearing

Auditory modeling of the hearing is motivated to contribute to the understanding of the hearing system, and to produce tools that mimic as accurately as possible the function that is being emulated [2]. Models of spatial hearing aim at emulating the functions of the auditory system that are related to the localisation of sounds. The inputs to the model are usually the acoustic signals that reach the two ears, and the direction of arrival is estimated based on, e.g., spectral features of the individual signals (monaural cues), or the differences in timing, level, and similarity between the two ears (binaural cues). When the location of a sound source is restricted only to the frontal horizontal plane, binaural cues are typically enough to resolve the direction of the sound. Thus, an auditory model that does not take advantage of the monaural cues suffices when the localisation task is limited to the frontal horizontal plane, as in the current study. Binaural auditory models can be found in *The Auditory Modeling Toolbox* by Søndergaard and Majdak [3] (i.e. [4–6]), and are included in the book *The technology of binaural listening*, by Blauert et al. [7].

Even though the auditory system is able to determine the direction of a sound source based on binaural cues, it is not clear how the details of that process play out. The timing and level differences between the two ears, interaural time differences (ITD) and interaural level differences (ILD), respectively, are frequency-dependent and may in some cases provide conflicting information in different frequency bands. Nevertheless, the cues are ultimately combined into a single perceived direction. The auditory nerve carries the signals from the cochlea to the cochlear nucleus, where the information is distributed to the superior olivary complex, which is divided into the medial superior olive (MSO) and the lateral superior olive (LSO). These two olivary complexes are known to be sensitive to ITD and ILD and appear to have a key role in azimuth localisation [8, 9], even though these binaural cues are not enough to resolve sound source direction due to what is often called the cone of confusion (regions where multiple points in the space share the same ITD and ILD values). However, the precise neural mechanisms that underlie the extraction of binaural cues remains unknown [10, 11]. Thus, even when dealing with the restricted case of horizontal plane localisation, the mapping from the binaural signals to the perceived direction of the auditory event is far from trivial.

One approach for estimating the perceived sound source direction from binaural cues involves probabilistic models, which have been proposed by [12, 13]. In [13], ITD and ILD cues were compared to a database of ITD and ILD values for known source locations, and the most probable sound source position was estimated from a posterior

probability distribution. This model could also incorporate prior information in the form of “knowledge maps” weighting the probability distribution. In [12], a binaural auditory front-end interface was followed by a Gaussian Mixture Model. Sound source direction was estimated by maximizing the likelihood of a given observation to a database for each cochlear frequency band. The final direction estimation across frequency bands was based on a probabilistic model.

2.2 Acoustics and evaluation of head-worn devices

Head-worn devices may partially or completely occlude the ears, leading to distortions in frequency and time domains. In the specific case of sound source localisation, this may lead to altered binaural and monaural cues, which in turn may deteriorate localisation accuracy. The amount of attenuation of a hearing protection device or the physical shape of a protection helmet play a significant role on the introduced effect and it is not evident how a device distorts the spatial sound perception without a subjective test evaluation.

The evaluation of the introduced effect on spatial localisation by HWDs is essential when they are worn in duties where good conservation of communication abilities and spatial perception of auditory surroundings is important. An American National Standard on *Methods for Measuring the Effect of Head-worn Devices on Directional Sound localization in the Horizontal Plane* [14] presents three alternative methods for evaluating the effect of a HWD on horizontal plane localisation. All methods are based on subjective listening experiments, where the task is to indicate the active loudspeaker in a circular horizontal array of loudspeakers. The number of loudspeakers in the array as well as the method of indicating the active loudspeaker differ between the alternative methods of the standard. The second method in it is of interest in this study, which is designed to measure localisation error in the horizontal plane with a high-resolution response metric.

Several studies about HWDs evaluation on localisation tasks can be found in the literature. Vause and Wesley Grantham [15] found that wearing earplugs and protective headgear increased horizontal localisation errors when compared to an open ears condition both in frontal and lateral directions, being more notable for the latter. Their results suggest that earplugs may disrupt interaural cues. Similar results were found in a following study on the effect of hearing protectors by Bolia et al. [16], where wearing earmuffs or earplugs increased mean azimuth localisation error by 5°. Simpson et al. [17] found that head movement patterns and search time for sound source localisation were affected by hearing protection devices. Their results suggest that localisation cues may be especially disrupted in all dimensions when wearing earplugs and earmuffs together. Zimpfer and Sarafian [18] studied front-back, up-down and left-right localisation errors when different HWDs were worn. They found significant differences on sound localisation performance which depended on the type of HWD, where active hearing protection devices induced more errors than passive devices.

Schepker et al. [19] and Denk et al. [20] investigated the sound quality of commercially available hear-through devices with subjective listening experiments. In [19], perceptual sound quality was evaluated. Their results suggest that the main factor in perceived sound quality was determined by how similar the transfer functions of the inserted device and the open ear canal were. The closer the perceptual characteristics were to the open ear, the better. In [20] large differences between devices were found; some devices conserved open ear properties well, while others introduced severe artefacts and destroyed the binaural cues. However, the direct implications of these artefacts were not reported. In the current study, the relationship between measured spatial cues artefacts and subjective behavior is investigated and utilised in objectively characterising HWDs.

The mapping between sound source location and the corresponding auditory spatial cues depend on the interplay between source content and individual ear, head, and upper body morphology. The auditory system exhibits long-term adaptation to the individual auditory cues [21], and is remarkably robust to a wide range of surrounding acoustic conditions and source content. Adaptation to altered binaural and monaural cues require sound exposure, training with feedback or explicit training [22]. This adaptation requires extended periods of time, especially if no feedback and no explicit training is allowed. In the current study, each HWD is worn for a limited time and no feedback or training is given in the test. Thus, it is assumed that subjects don't adapt to any of the spatial cue distortions caused by the HWDs, and that their task performance reflects the degradation of spatial cues due to the HWDs.

2.3 Neural networks on parameter estimation

Machine learning, in the context of artificial intelligence, is the field of computer science that aims at developing tools for solving tasks for which they were not explicitly programmed. A task for machine learning is to generate a model that is able to estimate the output values (labels) of a function that is unknown from a given set of input values (features). Supervised learning is the technique that uses a labeled training dataset to build the hypothesis to predict the output.

Artificial neural networks, or simply neural networks (NN), are computing systems that estimate labels following a structure inspired by biological neural connections. A NN is formed by interconnected nodes, also called neurons, originally based on the model of the perceptron [23]. These neurons are structured in layers, which transmit information from the input layer (formed by the input features) to the output layer (labels) through so-called hidden layers [24]. Neurons and their interconnections build the hypothesis by learning weights that optimize the estimation error on estimating the labels by processing the input features.

Machine learning techniques are of special interest when the analytical solution of a system is unknown. NNs in particular are useful for their ability to generate non-linear hypotheses and to adapt to a wide range of systems. Studies

involving NNs on sound direction of arrival have been proposed in the past. Chakrabarty and Habets [25] trained a convolutional NN able to estimate the direction of arrival of broadband noise, even in acoustic conditions that were not included in the training set. Adavanne et al. [26] used recurrent convolutional NNs to estimate the direction of arrival for multiple sources. The model was able to determine the number of sources and their direction in a two-dimensional space. However, the direction of arrival computed with these techniques is not necessarily representative of the human perception mechanisms and could outperform the human ability to localise sounds.

A neural network-based approach was proposed in [27], where the model was trained to evaluate localisation performance in the horizontal plane when HWDs were worn. The inputs in this model are: a single value to represent the ITD computed using IACC between directional impulse responses, a single value to represent the ILD computed as the difference of directional transfer function above 1.5 kHz, and a proximal stimulus that contains spectral information. The model is able to predict realistic performances when earplugs are worn, but the estimation error increases when earmuffs are assessed.

3 Methods

3.1 Head-worn devices

In this study, six different HWDs were investigated. The devices span a wide range of hearing devices and use cases (see Tab. 1): A and B were open-back headphones; C and D were hearing protection devices (HPDs); E and F were protection helmets. An open ears (OE) condition was also included in the study.

Device B was a pair of AKG K702 headphones, and A was a modified version of the same model [28]. The design of A was originally proposed as a DIY solution for augmented reality audio. The most relevant modifications in device A are openings cut on the front and back sides of each earpad [28].

Head-worn device C was a HPD that is meant to work as a communication headset (SAVOX Noise-COM 200) with noise-cancelling features together with a hear-through system using two microphones (located at left and right earmuffs). In condition C1, the default active features, including hear-through, were used. The sound was attenuated 5.9 dB and 7.9 dB at the left and right dummy head ear, respectively, compared to the open ears condition.

In C2, the balance setting of the active sound was set three steps towards the left ear. According to the manufacturer, this translates to a 4.5 dB attenuation in the right ear compared to the default settings, while the left ear was maintained as in the default settings. In condition C2, the sound was attenuated 5.9 dB and 13.7 dB at the left and right dummy head ear, respectively, compared to the open ears condition.

Device D was a Peltor Pro-Tac II, which is a HPD with similar characteristics to C. Default settings for device D

Table 1. Summary of head-worn devices.

ID	Type	Model	Settings
A	Headphones	AKG K-702	DIY version [28]
B	Headphones	AKG K-702	Original
C1	HPD	SAVOX Noise-COM 200	Default
C2	HPD	SAVOX Noise-COM 200	Balanced
D	HPD	Peltor Pro-Tac II	–
E1	Helmet	Bullard Magma	Visor down
E2	Helmet	Bullard Magma	Visor up
F	Helmet	Gecko MK11	–

were not specified in the user manual, and only the active hear-through gain was adjustable. The gain could be selected from no active sound present to a high level of active hear-through. The selected gain was set and maintained during the duration of the study and was the same for all participants and the HRTF measurements. In condition D, the sound was attenuated 12.9 dB and 14.5 dB at the left and right dummy head ear, respectively, compared to the open ears condition.

Head-worn device E was a Bullard Magma firefighter protection helmet, which was studied in two conditions: E1, with visor down, and E2, with visor up. Device F was a Gecko MK11 marine safety helmet, which has a pair of holes in the ears position and without a visor.

3.2 Subjective evaluation

3.2.1 Apparatus

The listening experiment was conducted in a variable-acoustics room at the Aalto Acoustics lab facilities ($RT60_{250\text{ Hz}} = 0.2363\text{ s}$, $RT60_{500\text{ Hz}} = 0.3272\text{ s}$, $RT60_{1\text{ kHz}} = 0.3421\text{ s}$, $RT60_{2\text{ kHz}} = 0.5284\text{ s}$, $RT60_{4\text{ kHz}} = 0.5671\text{ s}$, $RT60_{8\text{ kHz}} = 0.04819\text{ s}$). The experiment was designed to be similar to the second method of *ANSI/ASA S3.71-2019* [14]. For example, the loudspeaker array, the stimulus and the response grid were as recommended in the method. However, the procedure in the standard was modified in order to shorten the time of the experiment, so that a greater number of HWDs could be investigated. Thirty-six Genelec 1029A loudspeakers were mounted on a circular ring of 1.5 m radius, resulting in a 10° -spaced loudspeaker array. Only the loudspeakers at positions 0° , $\pm 10^\circ$, $\pm 30^\circ$, $\pm 50^\circ$, $\pm 70^\circ$ and $\pm 90^\circ$ were used for this study. The height of the array was adjusted so that each subject's ears were level with the loudspeakers.

Subjects sat on a rotatable chair, fixed in the center of the loudspeaker array. The chair was fitted with an adjustable head-rest to ensure correct positioning of the head before each trial.

The loudspeakers were obscured by a black, acoustically transparent curtain, so that subjects did not have information about the number and location of loudspeakers. Attached on the curtain, there was a numbered response grid with a 2° spacing between grid markers. Subjects gave their responses using a tablet computer, which showed the same response grid on a graphic user interface. Even though the response grid was 2° -spaced, the precision of

the user interface was higher, and answers were saved with an accuracy of 1° .

3.2.2 Listening experiment

Nineteen subjects (average age 28.68 years, 3 female/16 male) with self-reported normal hearing took part in the listening experiment. All subjects were employees at the Aalto Acoustics Lab and are considered experienced listeners. The study complies with the Declaration of Helsinki and was approved by the Research Ethics Committee of Aalto University. The participants provided written informed consent.

The total experiment consisted of 9 rounds, one round for each described condition wearing a HWD (A–F) and one without wearing any. Each round consisted of 27 trials: three repetitions were collected for each azimuth (0° , $\pm 10^\circ$, $\pm 30^\circ$, $\pm 50^\circ$ and $\pm 70^\circ$) in a randomized order, which was different for every round and every subject. The stimulus was a 250 ms pink noise burst, with an A-weighted level of 70 dB SPL, measured at the listener's position.

The chair was equipped with a head support which aided in head positioning. The subjects were instructed to place their head in the correct position before each trial using the head support as a reference, which was adjusted for each subject. After making sure the head was correctly positioned and facing 0° , the user pressed a button to start the trial. After a 1-second delay, the stimulus was presented, after which the subject was allowed to move their head. The task for the subject was to report the perceived direction of the sound source. After giving their answer with the tablet computer they pressed a button to confirm their response and move to the next trial. Before starting the next trial, they were again reminded to position their head correctly before they could start. The subjects were allowed to take as many breaks as desired and to follow their own pace.

Prior to the experiment, the subjects completed one full training round in the open ears condition to ensure that they understood the task and were familiar with the user interface.

3.3 HRTF measurements

HRTFs with the HWDs in place were measured with a G. R.A.S. KEMAR 45 BC head and torso simulator with anthropometric pinnae. The KEMAR was placed into the large anechoic chamber at Aalto University Acoustics Lab and mounted on a digitally-controlled turntable. The center

of the dummy head was at a distance of 1.5 m from a Genelec 8331 loudspeaker, which was used for playback of the measurement signals. The KEMAR ear simulators RA0045 (Shore 00-35) were connected to a G.R.A.S. 12 AG 8 channels power module, from which the signal was routed to the line input of a Fireface UFX+ sound interface.

The HWDs were fitted onto the G.R.A.S. KEMAR 45 BC dummy head. Exponential sweeps from 10 Hz to 25 kHz with a length of 1 s and an average presentation level of 80 dB SPL were presented at 1.5 m to record the impulse responses. Free-field reference responses were measured using a free-field G.R.A.S. Type 46AF microphone set located at the position of the center of the dummy head without the manikin being present, and were used for deriving the HRIR.

The KEMAR was mounted on a turntable at a 1.5 m distance from the loudspeaker. The turntable was automatically turned 10° between measurements, creating a 10° spaced set of measurements in the horizontal plane. Special attention was paid when fitting the HWDs onto the KEMAR to avoid possible errors due to their location. The measurements were checked in-situ to avoid unreliable recordings caused by misplacement of the HWDs. The measured impulse responses were corrected using the free-field microphone response to get the HRIRs [29].

3.4 Auditory model processing

A functional model of binaural hearing based on interaural cross-correlation was applied to compute binaural cues. The model is adapted from [10], based on the coincidence detector model proposed by Jeffress [30]. Although cross-correlation-based auditory modeling is not bolstered by neurophysiological evidence [8], the approach has been proven to explain directional perception in free field with good accuracy [2]. As in the subjective listening test, pink noise stimuli of 250 ms were used and convolved with the HRIRs from the measurements. Anechoic HRIRs were used, since we assume that the directional perception is based mostly on direct sound in the listening test. The early reflections and the reverberation time were relatively low in level in the room.

Peripheral frequency selectivity was modeled with a gammatone filterbank [31] from the Auditory Modeling Toolbox [3], in which center frequencies are distributed on the equivalent rectangular bandwidth (ERB) scale between 50 Hz and 8 kHz [32]. To mimic neural activity patterns, the bandpass filtered signals were half-wave rectified and low-pass filtered using a first-order IIR with a cut-off frequency of 1 kHz. The interaural cross-correlation approach was used to find out the maximum correlation value to estimate the ITD for each cochlear frequency band (adapted from [10]). This was computed within a maximum lag of 760 μ s at each time frame of 20 ms (hop size = 10 ms) and frequency band between the two ears. The level difference in dB SPL between the signals at each time frame was computed to get the ILD values for each cochlear frequency band.

ITD and ILD values were averaged over time to get an ITD and an ILD value for each listening test condition and

azimuth direction. Values of ITD up to 1.5 kHz and ILD from 1 kHz [2] were gathered as input features for the neural network model (see Fig. 1).

3.5 Estimation of perceived direction using neural networks

As mentioned before, the estimation of the direction perceived by a subject from the spatial cues measured from ear canal signals has been found to be a complicated process that depends on many factors, such as source direction and signal content. Nevertheless, in the case investigated in the present study, the task may be simpler. The aim of this study is to develop tools to evaluate HWDs as an alternative to running standardised listening tests, which narrows down the task quite prominently. Instead of explaining directional perception in general case, it is meaningful to simulate the perception only in the listening scenario of the listening test defined in the standard. The test uses short noise bursts in the horizontal plane, which potentially makes the machine learning task easier.

It should also be noted that the aim of the computational model is not to emulate human sound localisation behavior in detail, but to predict HWD-induced changes in localisation performance. Therefore, it is not the focus of the current study to evaluate whether the computational model results in a realistic representation of the human binaural system. The results and the computational model presented in the current study are not expected to generalise to other problem formulations involving binaural auditory processing.

Two artificial neural networks were trained using ITD and ILD estimates as input features and subjective data from the listening test as labels. Each training sample included 36 input features: 18 ITD values, one for each cochlear frequency band from 50 Hz to 1.5 kHz; and 18 ILD values, one for each cochlear frequency band from 1 kHz to 8 kHz. Both neural networks aimed at estimating data obtained from subjective tests.

For each subject response on perceived direction θ , average μ and standard deviation σ were computed:

$$\mu(\theta_{i,\varphi,s}) = \frac{1}{M_{i,\varphi,s}} \sum_{m=1}^{M_{i,\varphi,s}} \theta_{i,\varphi,s,m},$$

$$\sigma(\theta_{i,\varphi,s}) = \sqrt{\frac{1}{M_{i,\varphi,s}} \sum_{m=1}^{M_{i,\varphi,s}} [\theta_{i,\varphi,s,m} - \mu(\theta_{i,\varphi,s})]^2},$$

where:

- $\theta_{i,\varphi,s,m}$ is the perceived direction during HWD condition i , sound source direction φ , by subject s in trial m ;
- $M_{i,\varphi,s}$ is the number of trials for HWD condition i , sound source direction φ and subject s ;
- m is the trial.

One of the NNs aimed at estimating the perceived direction of the source. For each sound source direction and HWD condition, the average of responses over all subjects was computed:

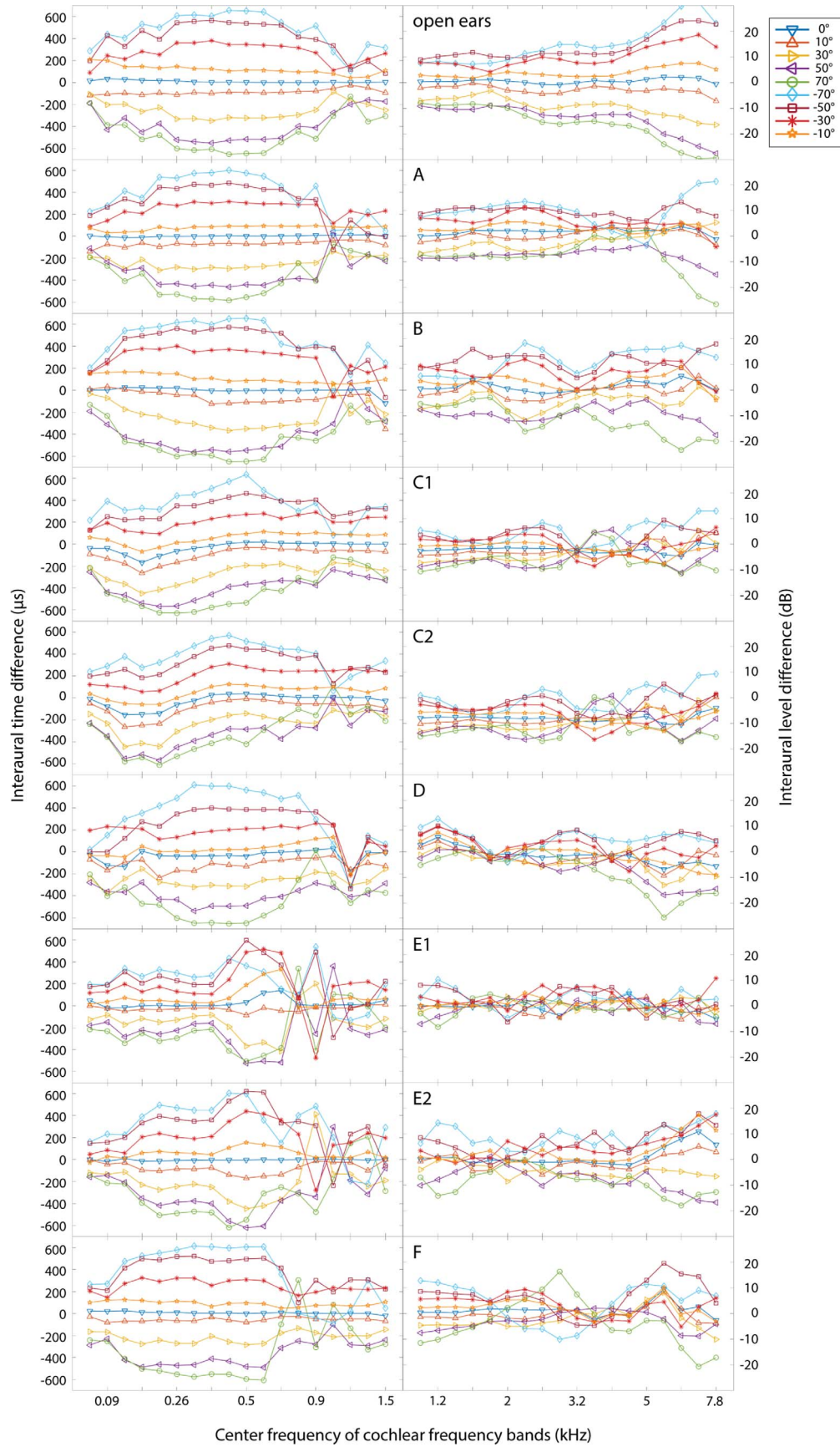


Figure 1. *Left:* ITD values per frequency band at the output of the auditory model. *Right:* ILD values per frequency band at the output of the auditory model.

$$\mu(\theta_{i,\varphi}) = \frac{1}{N_s} \sum_{s=1}^{N_s} \mu(\theta_{i,\varphi,s}),$$

where:

s is the subject;
 N_s is the number of subjects.

This perceived angle average, $\mu(\theta_{i,\varphi})$, was used as a label, $y_{\text{direction},i,\varphi}$, to train the NN to estimate the perceived direction of the source:

$$y_{\text{direction},i,\varphi} = \mu(\theta_{i,\varphi}).$$

The other NN aimed at estimating the position uncertainty of the source. We assumed in this study that the standard deviation of azimuth responses in the subjective test reflects the perceived position uncertainty of the sound source. For each sound source direction and listening condition, the standard deviation of responses among all subjects was computed:

$$\sigma(\theta_{i,\varphi}) = \sqrt{\frac{\sum_{s=1}^{N_s} \sigma(\theta_{i,\varphi,s})^2}{N_s}}.$$

This standard deviation, $\sigma(\theta_{i,\varphi})$, was used as a label, $y_{\text{uncertainty},i,\varphi}$, to train the NN to estimate the position uncertainty of the source:

$$y_{\text{uncertainty},i,\varphi} = \sigma(\theta_{i,\varphi}).$$

A data augmentation technique was used to improve the generalisation of the model. The size of the dataset was augmented 10 times by adding Gaussian noise to the input feature vectors, which was 60 dB SNR, empirically chosen to provide enough variation to the data not to overfit and, at the same time, maintain meaningful values when new data was generated. To achieve a better generalisation and avoid overfitting, each model was trained five times and the estimated outputs of each iteration were averaged to compute the NN-estimated output.

The amount of data was relatively small, which could potentially lead to results where the network would memorise each trained HWD case perfectly, making the result questionable in terms of generalisation to HWD models not used in training. To avoid this effect, a strategy based on the Leave-one-out (LOO) technique was used. LOO is a particular case of cross-validation algorithms that uses, for each iteration, the whole dataset as training data except one sample that is used as validation data. This process of training and evaluation is iterated for each sample until all samples are evaluated. For our particular case, a similar approach was used leaving one HWD out of the training data and using the excluded device data as test data. After all the iterations, all devices were evaluated. This approach aims to extrapolate how the NN would perform in a general case at estimating the effect of a HWD that has not been evaluated before.

To estimate the position uncertainty of the source when wearing a HWD, a single 16 neurons hidden layer fully

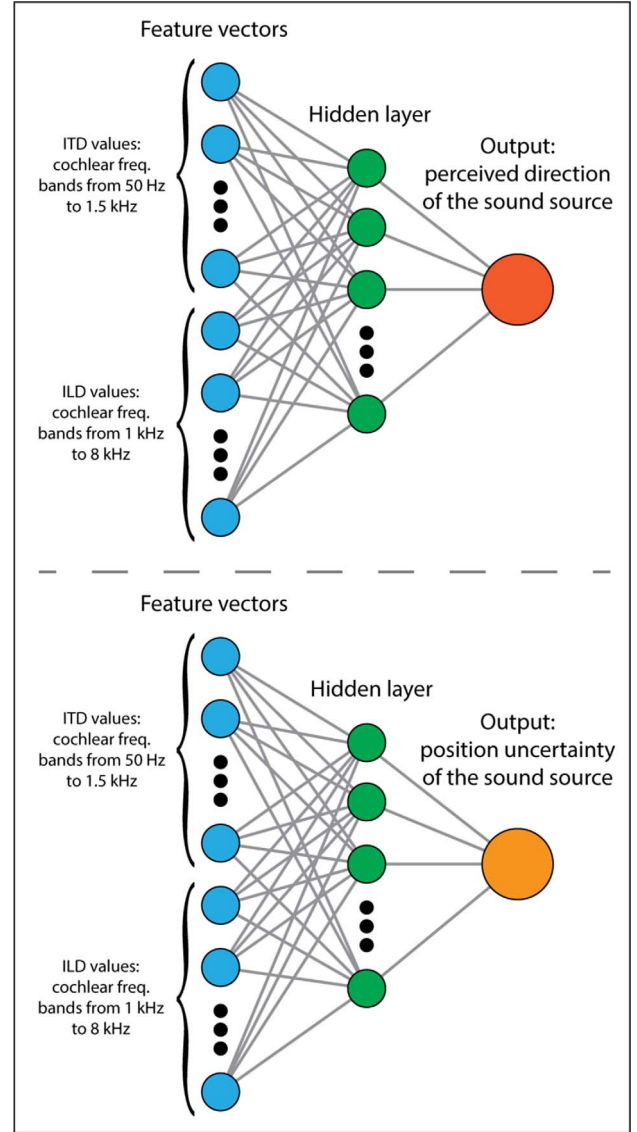


Figure 2. Neural Networks scheme. Two neural networks were trained to estimate perceived direction and position uncertainty of a sound source using ITD and ILD values computed with a binaural auditory model.

connected NN with Levenberg–Marquardt backpropagation was trained to minimise the mean squared estimation error by optimising the network parameters. The NN was trained with the rest of the HWD conditions, using the ITD and the ILD values as input features and $\sigma(\theta_{i,\varphi})$ as labels.

To estimate the perceived direction when wearing a HWD, a single 22 neurons hidden layer fully connected NN with Levenberg–Marquardt backpropagation was used to train the networks to minimise the mean squared estimation error by optimising the network parameters. The NN was trained with the rest of HWD conditions, using the ITD and ILD values as input features and $\mu(\theta_{i,\varphi})$ as labels (Fig. 2).

Table 2. Subjective experiment results. Average and standard deviation (in parenthesis) of perceived direction for each source azimuth angle and each listening condition.

Device φ	70°	50°	30°	10°	0°	-10°	-30°	-50°	-70°
OE	66.96° (5.04°)	45.07° (3.44°)	27.11° (4.05°)	9.19° (2.19°)	-0.31° (1.11°)	-8.74° (1.93°)	-28.28° (3.60°)	-44.56° (4.63°)	-67.43° (5.15°)
A	63.48° (5.51°)	41.07° (5.37°)	24.81° (3.36°)	8.31° (3.07°)	-0.50° (1.56°)	-8.88° (2.94°)	-26.74° (8.92°)	-39.74° (3.29°)	-65.61° (5.59°)
B	77.54° (4.06°)	53.31° (5.01°)	28.37° (3.44°)	10.37° (2.56°)	0.17° (1.40°)	-6.67° (2.63°)	-28.80° (6.22°)	-49.71° (7.83°)	-77.81° (5.20°)
C1	67.04° (6.33°)	42.74° (3.64°)	25.72° (4.11°)	7.15° (2.93°)	-0.78° (2.06°)	-8.80° (3.28°)	-32.65° (8.46°)	-49.91° (5.79°)	-74.69° (6.95°)
C2	71.30° (6.94°)	51.91° (7.65°)	26.63° (7.23°)	8.17° (2.54°)	1.59° (3.19°)	-6.46° (3.49°)	-32.11° (8.51°)	-53.02° (9.53°)	-76.43° (6.48°)
D	66.33° (6.36°)	49.72° (7.80°)	26.28° (5.90°)	7.31° (2.80°)	-0.78° (1.59°)	-9.63° (4.38°)	-29.37° (6.28°)	-54.00° (7.59°)	-75.30° (4.02°)
E1	65.96° (6.12°)	51.00° (5.63°)	28.00° (4.59°)	10.43° (2.34°)	-0.81° (2.49°)	-9.89° (3.70°)	-27.39° (4.61°)	-50.09° (4.57°)	-70.80° (7.09°)
E2	66.31° (6.51°)	43.67° (3.67°)	25.81° (2.02°)	8.93° (2.13°)	-0.13° (1.43°)	-8.85° (2.85°)	-27.06° (2.74°)	-40.85° (3.23°)	-67.43° (6.74°)
F	68.33° (6.07°)	45.57° (5.53°)	26.06° (2.40°)	8.69° (2.22°)	-0.81° (1.51°)	-9.81° (4.12°)	-29.57° (3.39°)	-46.13° (5.18°)	-69.91° (6.10°)

4 Results

4.1 Subjective evaluation results

For each participant, the average and standard deviation of responses per listening condition and sound source were computed. The data of subject 6 was discarded from analysis due to an abnormal strategy on responding when front-back confusions were encountered, responding always 90° when the sound was perceived from the back. Apart from this aforementioned subject, two trials were discarded from analysis, since it appeared clear that a mistake had been made using the user interface which had a high impact on the neural network performance. The first discarded trial was one of subject 8 in condition A and -10° source stimulus; responses were: -90°, -1°, -4°. As -90° was the default value if no response was given and it was the first trial of that round, the -90° response was excluded. The second discarded trial was one of subject 14 in condition C2 and 10° source stimulus; responses were: 0°, 0°, 89°. That response value generated a distortion in the standard deviation. Leaving it out provided a better representation of the results, so 89° response was excluded.

The average and standard deviation values of perceived azimuth for each listening condition and source angle are shown in Table 2. A one-way ANOVA was conducted for each source azimuth angle to compare the effect of HWDs on perceived direction (see Tab. 3). The effect of the HWD reached statistical significance ($p < 0.05$) at 70°, 50°, 0°, -10°, -50° and -70°. Post hoc t-tests with Bonferroni correction showed evidence of statistically significant differences in 17 out of 36 possible pairs of HWDs for at least in one of the studied source azimuth directions (see Tab. 4).

The HWD-induced angular error was computed as the absolute difference between the perceived azimuth using a HWD and the perceived azimuth in the open ears

Table 3. One-way ANOVA for each source azimuth angle (statistical significance is reached when $p < 0.05$).

φ	F	p
70°	$F(8, 153) = 2.65$	< 0.001
50°	$F(8, 153) = 3.75$	< 0.001
30°	$F(8, 153) = 0.86$	< 0.5551
10°	$F(8, 153) = 1.62$	< 0.1231
0°	$F(8, 153) = 2.68$	< 0.001
-10°	$F(8, 153) = 2.07$	< 0.0424
-30°	$F(8, 153) = 1.02$	< 0.42
-50°	$F(8, 153) = 3.95$	< 0.001
-70°	$F(8, 153) = 4.29$	< 0.001

Table 4. Summary of pairwise t-test with Bonferroni correction for all azimuth directions. (* $p < 0.05$ at least in one of the studied source azimuth directions).

Device	OE	A	B	C1	C2	D	E1	E2	F
OE	=	-	*	-	-	-	*	*	-
A	-	=	*	-	*	*	*	-	-
B	*	*	=	*	-	-	*	*	-
C1	-	-	*	=	*	-	-	-	-
C2	-	*	-	*	=	*	*	*	*
D	-	*	-	-	*	=	-	*	-
E1	*	*	*	-	*	-	=	-	*
E2	*	-	*	-	*	*	-	=	-
F	-	-	-	-	*	-	*	-	=

condition. As mentioned before, in this study position uncertainty of the source is assumed to be represented by the standard deviation of azimuth responses in the subjective test data. In Figure 3, position uncertainty of the source and HWD-induced angular error is shown. Correlation analysis was performed obtaining $r = 0.686$,

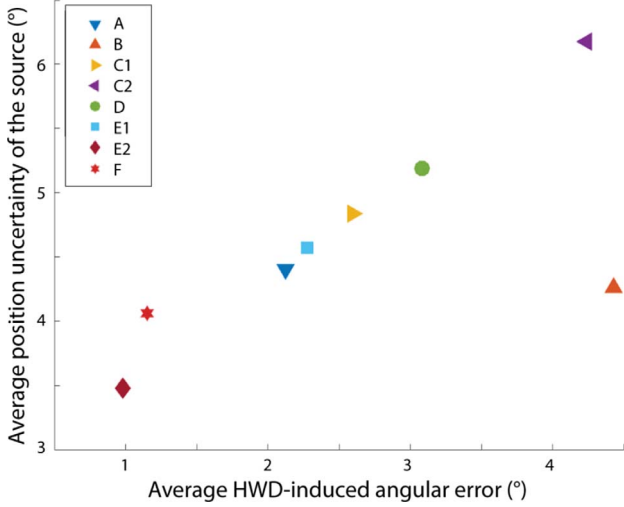


Figure 3. The relationship between average position uncertainty of the sound source and average HWD-induced error from the subjective test. The position uncertainty was computed as the standard deviation of the azimuth responses from the listening test data. The HWD-induced angular error was computed as the absolute difference between perceived direction using a HWD and perceived direction in the open ears condition.

$p = 0.0603$, which doesn't show statistical significance of correlation ($p < 0.05$). However, a linear dependency trend can be observed for 7 out of 8 devices.

4.2 Neural network model results

A NN predicted the HWD effect on position uncertainty of the source. For each HWD i , averages of the standard deviation of responses from the subjective test were computed, $y_{\text{uncertainty},i}$ and sorted to create a ranking of HWDs by position uncertainty. Similarly, averages of NN-estimated position uncertainty values were computed $\hat{y}_{\text{uncertainty},i}$. NN-estimated values were compared to actual subjective test values for each HWD, which resulted in a perfect position in the ranking for five HWDs: A, C1, C2, E2 and F; a relatively small error for two HWDs: B and D; and a large mismatch for HWD E1 (see Fig. 4).

Mean absolute error (MAE) was computed as the average prediction error generated by the NN when its output was compared to actual subjective test data. For NN-estimated values, average NN-estimation errors, $\text{MAE}_{\text{uncertainty},i}$ and $\text{MAE}_{\text{direction},i}$ over all samples for HWD i was defined as:

$$\text{MAE}_{k,i} = \frac{\sum_{n=1}^{N_i} |y_{k,i,n} - \hat{y}_{k,i,n}|}{N_i},$$

where:

- $\text{MAE}_{k,i}$ is the mean absolute error for estimating values of type k for HWD condition i ;
- k is uncertainty or direction, depending on the training labels;
- $y_{k,i,n}$ is the actual value of type k for HWD condition i on sample n ;

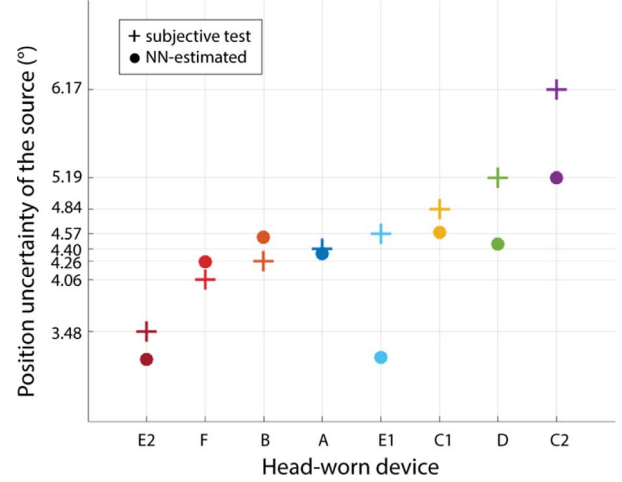


Figure 4. Comparison between position uncertainty of the sound source from the subjective test data and NN-estimated position uncertainty averages.

$\hat{y}_{k,i,n}$ is the NN-estimated value of type k for HWD condition i on sample n ;

N_i is the number of samples of HWD condition i .

The average estimation error of the NN on position uncertainty of the source for all HWDs together was $\text{MAE}_{\text{uncertainty,all}} = 1.22^\circ$. 6 out of 8 HWDs conditions were estimated with a $\text{MAE}_{\text{uncertainty},i} < 1.5^\circ$. NN-estimated data is compared to subjective data for each HWD in Figure 5.

The other NN predicted the HWD effect on perceived source direction. The average estimation error of the NN on perceived source direction for all HWDs together was $\text{MAE}_{\text{direction,all}} = 7.08^\circ$. 6 out of 8 HWDs conditions were estimated with a $\text{MAE}_{\text{direction},i} < 7^\circ$. NN-estimated data is compared to subjective data for each HWD in Figure 5.

The NN weights for each input feature were collected to analyse the relative importance of each binaural cue to the final decision. These learnt weights to estimate the subjective data are shown in Figure 6.

5 Discussion

The current study tested the effect of HWDs on the perceived direction and position uncertainty of a sound source in the horizontal plane. A listening condition without HWD was tested to serve as a baseline. The results in an open ears condition met the authors' expectations; perceived direction was biased towards the center and position uncertainty was higher for more lateral source positions. This higher position uncertainty is consistent with Blauert et al. [2], who reported that dependency between localisation blur and source azimuth angle was found from subjective data included in [33], which increases with sound source displacement from the forward direction. Bias on perceived direction may depend on the possibility of subjects to move their heads [34], with a tendency to overestimate azimuth angles when their head is fixed and to underestimate

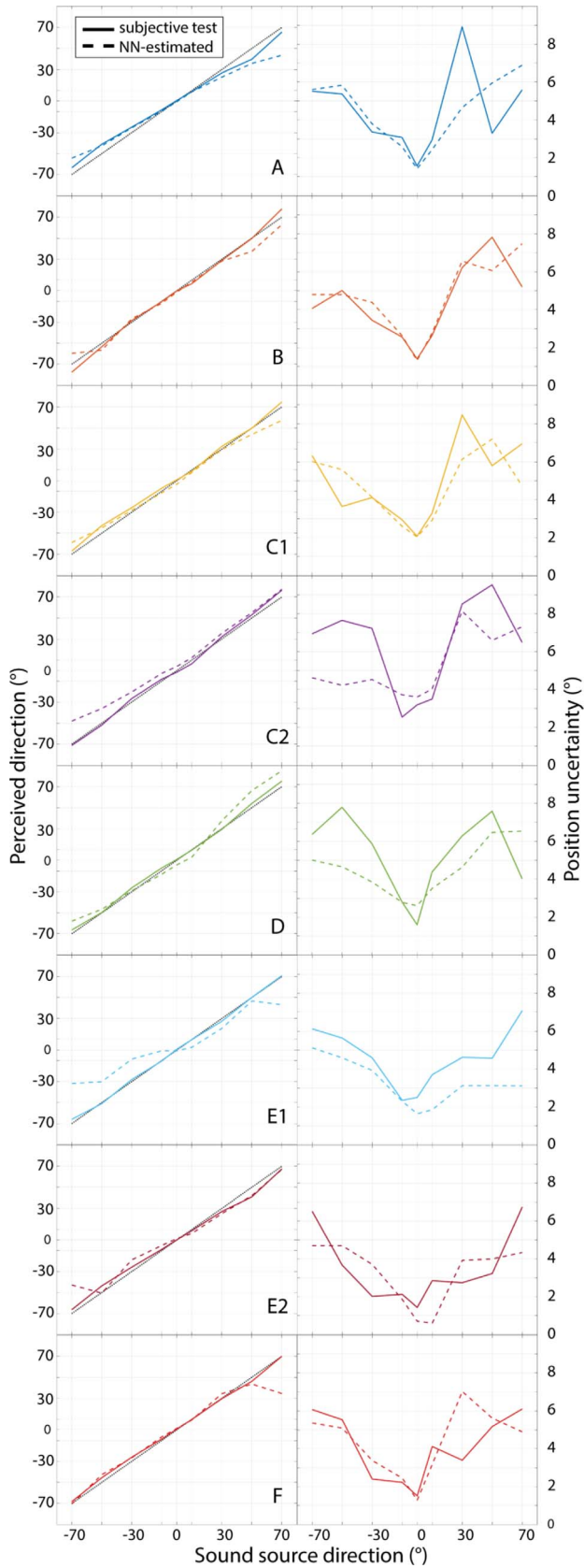


Figure 5. *Left:* perceived direction comparison between original subjective data and NN-estimated data in the horizontal plane. *Right:* position uncertainty comparison between original subjective data and NN-estimated data in the horizontal plane.

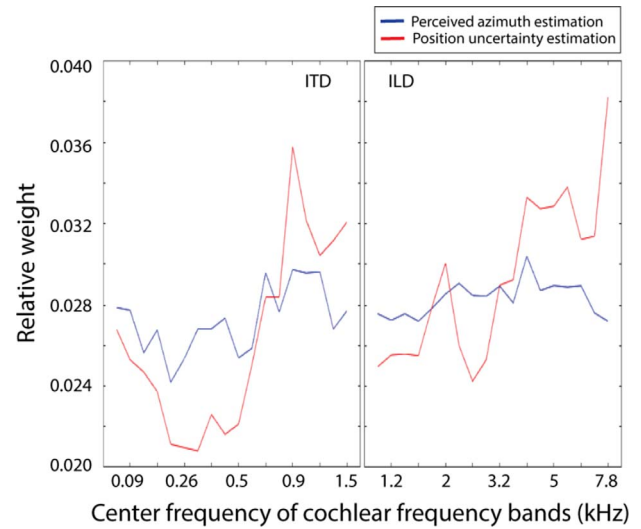


Figure 6. *Left:* ITD weights learnt by the NN to estimate subjective data for each cochlear frequency band. *Right:* ILD weights learnt by the NN to estimate subjective data for each cochlear frequency band.

it when subjects can move freely during the stimulus or instructed to face the source after the stimulus, which was the case for this study.

The results of the subjective test showed statistically significant differences in perceived direction in 6 out of 9 studied angles for the tested HWDs. These results are in agreement with previous studies. Bolia et al. [16] found significant differences on perceived direction in the horizontal plane when hearing protectors devices were tested. In their study, wearing hearing protectors occasioned an increase in the mean azimuth error on the order of 5° , which is in accordance with the results of this study. Similarly, Vause and Wesley Grantham [15] found a significant increase on localisation errors caused by the effect of earplugs and protective headgear.

The introduced effect of HWDs only reached statistical significance for conditions B, E1 and E2 when compared to the open ears condition. However, noticeable differences can be found both in azimuth average and standard deviation that may cause a practical effect even though significance wasn't reached. Differences in azimuth average and in standard deviation to the order of 5° may not seem enough to disturb localisation of a source in real scenarios, but these effects could lead to confusions and discomfort that would make a difference depending on the task they are worn for.

Comparison among HWDs wasn't straightforward due to inconsistency of the effect of the devices. For example, a device could conserve the perception of the source direction and position uncertainty for a tested source azimuth direction but affect both source direction and position uncertainty for another tested direction. For that, the overall performance was analysed by comparing the position uncertainty average of the sound among HWDs, represented in this study as the standard deviation of azimuth responses in the subjective test. Correlation between average position uncertainty of the source and average

HWD-induced direction error over azimuth angles wasn't statistically significant. Nonetheless, its dependency follows a linear trend for 7 out of 8 HWDs, which suggests that estimating the sound source position uncertainty using the standard deviation of azimuth responses from subjective test may be useful to predict the HWD-induced angular error.

Methods that objectively measure and quantify the effect introduced by a HWD can be found in the literature, especially in the context of hear-through devices and transparency evaluation. In [35], headphone influence was measured as the difference transfer function between the reference HRIRs (open ears) and the obstructed HRIR (when headphones were worn). A subjective test was conducted that revealed that change in perceived sound source position was the second most important feature for the participants to rate the devices transparency, after coloration introduction. Although the used objective measure may be a good estimator for coloration introduction, it is likely not enough to explain spatial effects without further analysis.

In [20], HRIRs were analysed and a binaural cues analysis based on interaural cross-correlation was conducted to examine distortions in ITD, ILD and interaural coherence. Peripheral processing was applied using third-octave bands around 250 Hz, 1 kHz, 4 kHz and 8 kHz to approximate auditory filters. Differences in ITD and ILD were found that depended on the hearing device and on the studied frequency band, which is in line with this study. Although devices effects were assessed subjectively in [19], no specific task about source localisation was conducted. Moreover, the relationship between objective data and perceptual evaluation was not addressed, and therefore it is not clear if the objective metrics were sufficient to explain the effect of the analysed HWDs.

The computational model proposed in this study utilises the output of an auditory model as the input for a NN that estimates HWD-induced effects in sound source localisation. This approach allows the usage of the state of the art auditory models for application oriented evaluations, and also enables the possibility of correlating objective and subjective data to assess the reliability of the used objective measures.

An auditory model-based method was built to estimate the effect of HWDs on horizontal localisation. At the output of the auditory model, ITD and ILD distortions due to different HWDs were estimated. As a general overview, ILD was significantly affected by all HWDs, while ITD was better conserved. This confirms that it is possible to estimate the effect of HWDs on horizontal localisation using an auditory model. In the results from [20], ILD was generally better conserved than ITD, which differs from the results of this study, at least for conditions C1 – F. Nonetheless, HWDs A and B could be comparable to the ear-worn devices included in [20], aimed at evaluating acoustic transparency, while HWDs C1 – F distorted severely the interaural cues, especially ILDs.

A NN approach was used to explore to prove the feasibility of estimating subjective evaluation data by machine learning, when actual subjective evaluation data is not

available. To keep the model as accurate as possible to the human hearing system, only the parts of the hearing system that are not fully understood were modeled with a NN. The learnt weights by the NN showed that all the included ITD and ILD frequency bands were used combined to estimate the subjective data. This is consistent with the auditory perception literature, being azimuth localisation dependent on a complex combination of binaural cues that remains unclear.

The relationship between the position uncertainty of a source (represented in this study as the standard deviation of azimuth responses in the subjective test) and HWD-induced error suggests that estimating the average position uncertainty would be useful to predict the average HWD-induced angular error. Ranking a new device among the training database using the aforementioned position uncertainty was successful, which suggests that it is possible to build a model that predicts the effect of a HWD on horizontal plane localisation. The NN-estimation error when predicting the perceived direction was kept reasonably low, since the average error was in the order of the subjective test data standard deviation of azimuth responses.

6 Summary and conclusion

A method to estimate the effect of head-worn devices on frontal horizontal localisation is proposed in this study. A subjective listening test to evaluate localisation performance in the horizontal plane when wearing head-worn devices is conducted, the results of which show significant differences on perceived azimuth direction when head-worn devices are worn. A binaural auditory model is used to compute ITD and ILD values for each cochlear frequency band, which are used as feature vectors for a neural network model that estimates the perceived direction and position uncertainty of a source from listening test data.

It is shown that the effect of head-worn devices on localisation ability of the wearer measured with subjective testing can be estimated using a computational method that utilises a simple model of binaural interaction and a machine learning stage implemented with neural networks. The level of degradation of perception of direction imposed by a head-worn device can be predicted. Future work could expand the scope of the method to evaluate other dimensions of spatial hearing, such as front-back confusions, elevation or distance.

Conflict of interest

Authors declared no conflict of interests.

Data availability statement

The model (llado2022) is publicly available as part of the Auditory Modelling Toolbox (AMT, <https://www.amtoolbox.org>) [36] in the release of the version 1.0.0 available as a full package for download [37]. The data and the implementation of the simulations are available for download in Zenodo (<https://doi.org/10.5281/zenodo.5785910> [38]).

Acknowledgments

The authors thank Savox Communications Oy Ab (Ltd) and Ilkka Huhtakallio for collaboration. This work was part of the HUMAN Optimized xR (HUMOR) Co-innovation project, funded by Business Finland.

References

1. J. Blauert, J. Braasch, J. Buchholz, H. Steven Colburn, U. Jekosch, A. Kohlrausch, J. Mourjopoulos, V. Pulkki, A. Raake: Aural assessment by means of binaural algorithms – the AABBA project, in Proceedings of the International Symposium on Auditory and Audiological Research, Vol. 2. 2009, pp. 113–124.
2. J. Blauert: Spatial hearing: The psychophysics of human sound localization. MIT Press, 1997.
3. P.L. Søndergaard, P. Majdak: The auditory modeling toolbox, in The technology of binaural listening, Springer. 2013, pp. 33–56.
4. W. Lindemann: Extension of a binaural cross-correlation model by contralateral inhibition. I. Simulation of lateralization for stationary signals. The Journal of the Acoustical Society of America 80, 6 (1986) 1608–1622.
5. M. Dietz, S.D. Ewert, V. Hohmann: Auditory model based direction estimation of concurrent speakers from binaural signals. Speech Communication 53, 5 (2011) 592–605.
6. T. May, S. van de Par, A. Kohlrausch: Binaural localization and detection of speakers in complex acoustic scenes, in The technology of binaural listening, Springer. 2013, pp. 397–425.
7. J. Blauert: The technology of binaural listening. Springer, 2013.
8. B. Grothe: New roles for synaptic inhibition in sound localization. Nature Reviews Neuroscience 4, 7 (2003) 540–550.
9. D.J. Tollin: The lateral superior olive: A functional role in sound source localization. The Neuroscientist 9, 2 (2003) 127–143.
10. V. Pulkki, M. Karjalainen: Communication acoustics: An introduction to speech, audio and psychoacoustics. John Wiley & Sons, 2015.
11. T. Hirvonen, V. Pulkki: Perception and analysis of selected auditory events with frequency-dependent directions. Journal of the Audio Engineering Society 54, 9 (2006) 803–814.
12. 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.
13. 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.
14. ANSI/ASA: Methods for measuring the effect of head-worn devices on directional sound localization in the horizontal plane, in ANSI/ASA S3.71, 2019 Edition. 2019.
15. N.L. Vause, D. Wesley Grantham: Effects of earplugs and protective headgear on auditory localization ability in the horizontal plane. Human Factors 41, 2 (1999) 282–294.
16. R.S. Bolia, W.R. D'Angelo, P.J. Misher, L.J. Morris: Effects of hearing protectors on auditory localization in azimuth and elevation. Human Factors 43, 1 (2001) 122–128.
17. B.D. Simpson, R.S. Bolia, R.L. McKinley, D.S. Brungart: The impact of hearing protection on sound localization and orienting behavior. Human Factors 47, 1 (2005) 188–198.
18. V. Zimpfer, D. Sarafian: Impact of hearing protection devices on sound localization performance. Frontiers in Neuroscience 8 (2014) 135.
19. H. Schepker, F. Denk, B. Kollmeier, S. Doclo: Acoustic transparency in hearable – perceptual sound quality evaluations. Journal of the Audio Engineering Society 68, 7/8 (2020) 495–507.
20. F. Denk, H. Schepker, S. Doclo, B. Kollmeier: Acoustic transparency in hearables – technical evaluation. Journal of the Audio Engineering Society 68, 7/8 (2020) 508–521.
21. J.P. Rauschecker: Auditory cortical plasticity: A comparison with other sensory systems. Trends in Neurosciences 22, 2 (1999) 74–80.
22. C. Mendonça: A review on auditory space adaptations to altered head-related cues. Frontiers in Neuroscience 8 (2014) 219.
23. F. Rosenblatt: The perceptron: A probabilistic model for information storage and organization in the brain. Psychological Review 65, 6 (1958) 386.
24. A.G. Ivakhnenko: Polynomial theory of complex systems. IEEE Transactions on Systems, Man, and Cybernetics 4 (1971) 364–378.
25. S. Chakrabarty, E.A.P. Habets: Broadband doa estimation using convolutional neural networks trained with noise signals, in 2017 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), IEEE. 2017, pp. 136–140.
26. S. Adavanne, A. Politis, T. Virtanen: Direction of arrival estimation for multiple sound sources using convolutional recurrent neural network, in 2018 26th European Signal Processing Conference (EUSIPCO), IEEE. 2018, pp. 1462–1466.
27. T. Joubaud, V. Zimpfer, A. Garcia, C. Langrenne: Sound localization models as evaluation tools for tactical communication and protective systems. The Journal of the Acoustical Society of America 141, 4 (2017) 2637–2649.
28. N. Meyer-Kahlen, D. Rudrich, M. Brandner, S. Wirler, S. Windtner, M. Frank: Diy modifications for acoustically transparent headphones, in Audio Engineering Society Convention 148, Audio Engineering Society. 2020.
29. M. Zhang, W. Zhang, R.A. Kennedy, T.D. Abhayapala: HRTF measurement on KEMAR manikin, in Annual Conference of the Australian Acoustical Society 2009 – Acoustics 2009: Research to Consulting, 23–25 November 2009, Adelaide, Australia. 2009.
30. L.A. Jeffress: A place theory of sound localization. Journal of Comparative and Physiological Psychology 41, 1 (1948) 35.
31. R.D. Patterson: The sound of a sinusoid: Spectral models. The Journal of the Acoustical Society of America 96, 3 (1994) 1409–1418.
32. B.R. Glasberg, B.C.J. Moore: Derivation of auditory filter shapes from notched-noise data. Hearing Research 47, 1–2 (1990) 103–138.
33. P.H. Schmidt, A.H. Van Gemert, R.J. De Vries, J.W. Duyff: Binaural thresholds for azimuth difference. Acta Physiologica et Pharmacologica Neerlandica 3, 1 (1953) 2–18.
34. J. Lewald, W.H. Ehrenstein: Auditory-visual spatial integration: A new psychophysical approach using laser pointing to acoustic targets. The Journal of the Acoustical Society of America 104, 3 (1998) 1586–1597.
35. C. Schneiderwind, A. Neidhardt, D. Meyer: Comparing the effect of different open headphone models on the perception of a real sound source, in Audio Engineering Society Convention 150, Audio Engineering Society. 2021.
36. P. Majdak, C. Hollomey, R. Baumgartner: Amt 1.0: The toolbox for reproducible research in auditory modeling. Submitted to Acta Acustica (2021).

37. The AMT team: The auditory modelling toolbox full package (version 1.x) [code]. 2021. <https://sourceforge.net/projects/amtoolbox/files/AMT%201.x/amtoolbox-full-1.0.0.zip/download>.
38. P. Lladó, P. Hyvärinen, V. Pulkki: Auditory model-based estimation of the effect of head-worn devices on frontal horizontal localisation (version 1.0) [code]. 2022. <https://doi.org/10.5281/zenodo.5785910>.

Cite this article as: Lladó P. Hyvärinen P. & Pulkki V. 2022. Auditory model-based estimation of the effect of head-worn devices on frontal horizontal localisation. Acta Acustica, 6, 1.