

Sensory and Motor Systems

# Adaptive Response Behavior in the Pursuit of Unpredictably Moving Sounds

José A. García-Uceda Calvo,  Marc M. van Wanrooij, and  A. John Van Opstal<https://doi.org/10.1523/ENEURO.0556-20.2021>

Donders Centre for Neuroscience, Department of Biophysics, Radboud University, Nijmegen 6525 AJ, The Netherlands

## Abstract

Although moving sound-sources abound in natural auditory scenes, it is not clear how the human brain processes auditory motion. Previous studies have indicated that, although ocular localization responses to stationary sounds are quite accurate, ocular smooth pursuit of moving sounds is very poor. We here demonstrate that human subjects faithfully track a sound's unpredictable movements in the horizontal plane with smooth-pursuit responses of the head. Our analysis revealed that the stimulus–response relation was well described by an under-damped passive, second-order low-pass filter in series with an idiosyncratic, fixed, pure delay. The model contained only two free parameters: the system's damping coefficient, and its central (resonance) frequency. We found that the latter remained constant at  $\sim 0.6$  Hz throughout the experiment for all subjects. Interestingly, the damping coefficient systematically increased with trial number, suggesting the presence of an adaptive mechanism in the auditory pursuit system (APS). This mechanism functions even for unpredictable sound-motion trajectories endowed with fixed, but covert, frequency characteristics in open-loop tracking conditions. We conjecture that the APS optimizes a trade-off between response speed and effort. Taken together, our data support the existence of a pursuit system for auditory head-tracking, which would suggest the presence of a neural representation of a spatial auditory fovea (AF).

*Key words:* auditory fovea; auditory motion perception; head movement; human; linear systems; sound localization

## Significance Statement

Inspired by the visual ocular smooth-pursuit system, several studies have used eye movements to track moving sounds, but obtained poor pursuit performance, which led to the idea that the auditory system lacks sensitivity to sound velocity. We here demonstrate accurate head-pursuit of sounds, moving along unpredictable trajectories in the horizontal plane. Interestingly, the auditory pursuit responses adapted to the covert movement spectrum of the stimulus ensemble, from which we infer that the system may optimize a trade-off between movement speed and effort. Our results support the existence of an auditory pursuit system (APS), and we discuss its implications for the neural mechanisms that represent and track moving sounds.

## Introduction

To infer source directions in the horizontal plane of the head, the auditory system extracts interaural differences in arrival time and sound level [interaural level differences (ITDs) and interaural timing differences (ILDs),

respectively; Middlebrooks and Green, 1991; Blauert, 1997]. Front-back and up-down localization relies on the interaction of sound-waves within the pinnae, resulting in idiosyncratic direction-dependent spectral acoustic filters (Musicant and Butler, 1984; Oldfield and Parker, 1984; Wightman and

Received December 22, 2020; accepted March 13, 2021; First published April 19, 2021.

The authors declare no competing financial interests.

Author contributions: M.M.v.W. and A.J.V.O. designed research; J.A.G.-U.C. performed research; J.A.G.-U.C., M.M.v.W., and A.J.V.O. analyzed data; J.A.G.-U.C., M.M.v.W., and A.J.V.O. wrote the paper.

Kistler, 1989; Middlebrooks, 1992; Hofman et al., 1998; Van Wanrooij and Van Opstal, 2005).

However, auditory scenes typically contain moving sounds, and subjects may move actively or passively through the environment. As accurate sound-motion perception would enable the prediction of sound-source trajectories in the environment (Crum and Hafter, 2008), neural processing of dynamic acoustic-cue changes is crucial to track moving sounds (Vliegen et al., 2004).

Perceptual sensitivity to acoustic motion has been quantified by the minimum audible movement angle (Mills, 1958; Harris and Sergeant, 1971; Grantham, 1986). Sound-motion perception has typically been studied with the head stationary, inspired by studies of visual-motion mechanisms. An unresolved issue is whether moving sounds are processed by neural mechanisms tuned to continuous motion, or by a snapshot position-localization mechanism.

Moving visual targets are tracked with smooth-pursuit eye movements (Rashbass, 1961; Robinson, 1965; Krauzlis and Lisberger, 1994; Krauzlis, 2004; Barnes, 2008; Lisberger, 2010). Visual feedback provides the positional error and retinal slip velocity, needed to realign the fovea with the target through corrective saccades and smooth pursuit. Because of significant visual-motor delays ( $\approx 80$  ms; Robinson, 1965), visual feedback alone is insufficient for accurate pursuit, which also incorporates higher-level predictive mechanisms (Barnes, 2008). Neurons in visual-cortical motion areas like MST encode the direction and velocity of foveal stimuli and underlie the generation of accurate smooth-pursuit eye movements (Mikami et al., 1986; Dürsteler et al., 1987; Krauzlis and Lisberger, 1991; Ilg and Thier, 2008).

The question whether similar mechanisms exist in the auditory system has so far received little attention. Neurons in inferior colliculus (IC) and medial geniculate nucleus (cat: Al'tman et al., 1985; Bekhterev, 2003; guinea pig: Ingham et al., 2001; bat: Olsen and Suga, 1991; Pollack, 2012; barn owl: Wagner and Takahashi, 1992), and in auditory cortex (human EEG: Kreitewolf et al., 2011; cat: Stumpf et al., 1992; Toronchuk et al., 1992; Poirier et al., 1997; rat: Doan and Saunders, 2003; bat: Firzlaff and Schuller, 2001), have been shown to be sensitive to simulated dichotic sound motion. However, there is no conclusive evidence yet for an active auditory pursuit mechanism.

European Union program FP7-People-2013-ITN, European Union Horizon 2020 ERC Advanced Grant.

This work was supported by the European Union Program FP7-PEOPLE-2013-ITN "HealthPAC" 604063 (to J.A.G.-U.C. and M.M.v.W.) and European Union Horizon 2020 ERC Advanced Grant-2016 "Orient" 693400 (to A.J.V.O.).

Acknowledgements: We thank the valuable technical assistance of Günter Windau, Ruurd Lof, and Stijn Martens. We also thank the volunteers who participated in the experiments.

Correspondence should be addressed to A. John Van Opstal at [j.vanopstal@donders.ru.nl](mailto:j.vanopstal@donders.ru.nl).

<https://doi.org/10.1523/ENEURO.0556-20.2021>

Copyright © 2021 García-Uceda Calvo et al.

This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license, which permits unrestricted use, distribution and reproduction in any medium provided that the original work is properly attributed.

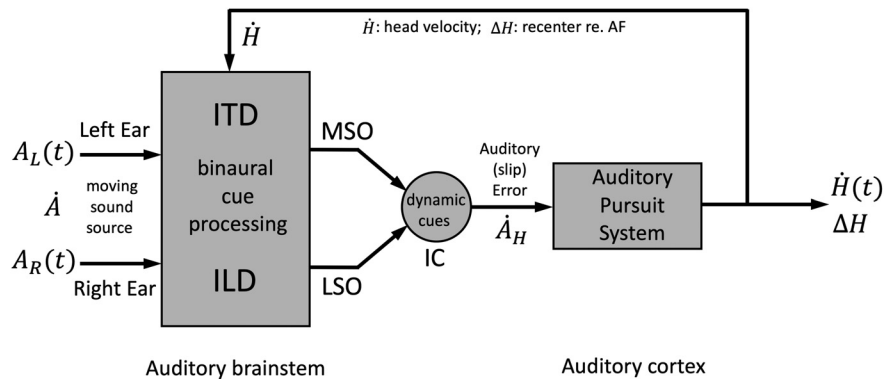
Brief sounds can elicit accurate goal-directed eye movements (Hofman and Van Opstal, 1998). Yet, smooth eye movements to moving sounds are practically non-existent, as ocular sound-tracking occurs through a series of saccades, with at best low-gain smooth-pursuit (Boucher et al., 2004; Berryhill et al., 2006). This led to the hypothesis that sound motion is perceptually tracked by intermittent sampling of the source position (Middlebrooks, 2015; Carlile and Leung, 2016), rather than by a continuous measurement of sound velocity.

However, as ocular sound-tracking does not affect any acoustic localization or pursuit error, we reasoned that, instead, an appropriate head-tracking response could keep craniocentric acoustic cues near the region of highest spatial acuity, just like in visual pursuit. So far, only few studies have investigated auditory-evoked head-tracking. For example, cats tracked apparent-motion clicks through multiple head-saccades, as may be expected from a snapshot mechanism (Beitel, 1999; Middlebrooks, 2015), and human head-tracking in a virtual-reality setup indicated that tracking accuracy degraded at higher simulated sound-velocities (Carlile and Leung, 2016).

Self-generated head movements facilitate the externalization of virtual-reality sounds (Brimijoin et al., 2013), which suggests a tight integration of the sound-localization cues with neural motor commands, and emphasizes the importance of sensorimotor integration in sound localization (Makous and Middlebrooks, 1990; Vliegen et al., 2004; Pallus and Freedman, 2016; Van Opstal, 2016). In line with this notion, it was recently demonstrated that active head movements significantly improve acoustic distance perception (Genzel et al., 2018). Such a sensorimotor relationship has so far not been studied for auditory pursuit under free-field hearing conditions.

In this article, we therefore characterized human head-movement pursuit to a free-field sound, moving along unpredictable trajectories in the horizontal plane. Listeners only had access to the acoustic input, and their self-generated head movements.

The rationale of our study is illustrated by the scheme in Figure 1. We hypothesized that, like for visual pursuit, accurate head tracking of the sound source requires an auditory pursuit system (APS) that would be driven by an auditory slip error. This error arises, because of an ongoing difference in sound and head velocity, and because the head-centered sound-location may differ with respect to a head-fixed auditory fovea (AF). The AF would represent the region of highest spatial acuity, and is presumably located around the straight-ahead direction, where the ILDs and ITDs are close to zero and have their highest resolution (Mills, 1958). Note that in contrast to the visual fovea in the retinae of both eyes, the representation of an AF would result from a neuro-computational mechanism, as it results from binaural integration. The auditory slip-error with respect to the AF,  $A_{\mu}(t)$ , results from the difference between sound velocity relative to the head,  $\dot{A}$ , and head velocity,  $H$ , and the localization error,  $\Delta H$ , which may all be derived from the dynamic changes in acoustic ITD/ILD cues in the auditory midbrain IC. A recentering (saccadic) head movement,  $\Delta H$ , would bring the sound



**Figure 1.** Presumed processing stages for horizontal auditory head pursuit. Both ears receive time-varying ILD and ITD because of a moving sound in the horizontal plane, and the head turning at angular velocity  $\dot{H}$ . Integration of these dynamic cues provides an estimate of sound velocity with respect to the head,  $\dot{A}_H(t)$  (auditory slip velocity). The APS aims to minimize the auditory slip-velocity error and the source-position error,  $\Delta H$ , with respect to the Auditory Fovea (AF), by bringing and keeping the instantaneous ILD and ITD cues close to zero.

close to the AF, so that when head velocity and position equal sound velocity and position, medial superior olive (MSO) and lateral superior olive (LSO) will both signal a (near-)zero ILD/ITD.

## Materials and Methods

### Subjects

Eleven subjects (S1–S11; five females; ages 21–43 years) participated in the experiments after providing their informed consent. All subjects had normal binaural hearing, had no motor problems, and normal or corrected-to-normal vision. The first author of this study was one of the participants. All other subjects were not aware of the purpose of the study. Five subjects had participated in other sound-localization experiments in the laboratory. To get familiarized with the experimental procedures, the naive subjects first received a short practice session before the actual experiments.

### Ethics

The experiments fully adhered to the protocols regarding observational experiments on healthy human adults and were approved by the local institutional ethical committee of the Faculty of Social Sciences (ECSW 2016-2208-41). All participants signed an informed consent form, before the start of the experimental sessions.

### Experimental setup

Subjects were seated in a completely dark anechoic chamber ( $3 \times 3 \times 3 \text{ m}^3$ ) in which the background noise level was  $\sim 30$ -dB SPL (A-weighted). Reflections above 500 Hz were effectively absorbed by black radio-absorbent material (UXEM Flexible Foams) that was mounted on the floor, walls, ceiling and on every large object present in the room.

The auditory stimuli were presented from a broadband loudspeaker (SC5.9, Visaton; Art. No. 8006) mounted on a custom-made L-shaped robotic arm that was driven by a DC motor (JVL MAC140-A1 integrated servomotor,

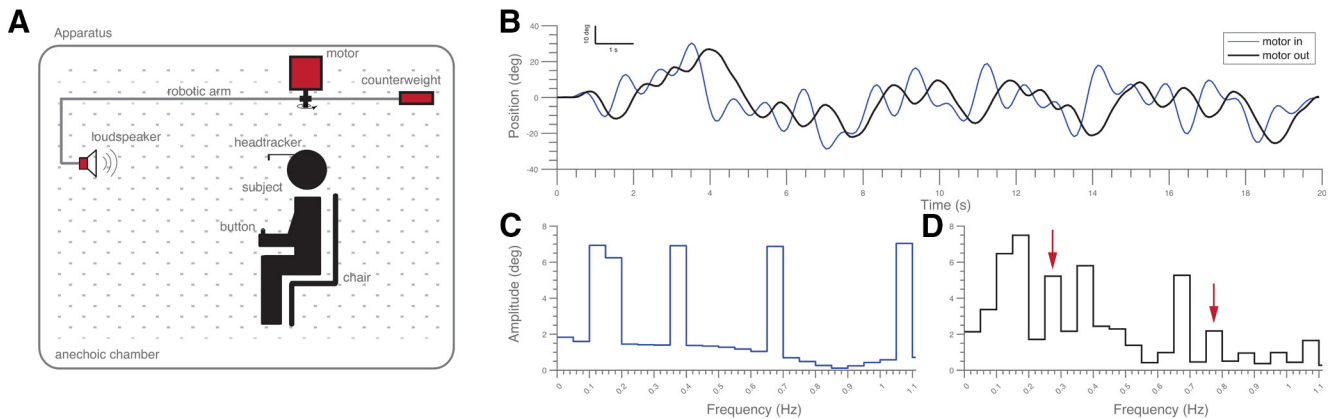
Gearbox Wittenstein Alpha–Angular Hollow Shaft MF2-50-5B1). The input signal for the motor was programmed in MATLAB (The MathWorks) and sent to the DC motor through a Tucker Davis Technologies TDT-RP2.1 ADC module. This setup (Fig. 2A) enabled rapid and accurate positioning of the speaker at a fixed distance of 1.15 m at any azimuthal direction around the subjects' head.

Head orientation and actual speaker movements were measured with a magnetic search-coil system (Robinson, 1963). Briefly, three orthogonal magnetic fields were generated by alternating currents of three different frequencies passing through three pairs of  $3 \times 3 \text{ m}^2$  squared coils, spanned along the edges of the room, which in turn induced alternating voltages in a small search coil (diameter,  $\sim 5 \text{ cm}$ ) attached on a light-weight glasses frame that was adjusted to fit on the subjects' head, without interfering with the ears. This system enabled accurate recording of 3D head orientations at a resolution of  $0.1^\circ$ , or better (Van Bentum et al., 2017).

From the center of the glasses frame, a 40 cm long, thin, aluminum rod (weight  $\sim 50 \text{ g}$ ) protruded forward with a small  $1\text{-cm}^2$  black plate attached to its end, which was positioned in front of the subjects' eye, and on which a dim red laser spot was projected from the subject's nose bridge. The laser spot served as a head-fixed and eye-fixed pointer, and helped the subject to fixate the eye-in-head orientation, while pointing with the head to the sound source. This procedure thus ensured accurate measurement of sound-evoked head movements without the co-occurring saccadic eye movements of natural gaze shifts.

### Stimulus characteristics

Auditory stimuli consisted of Gaussian white noise with a duration of 20 s. Sounds had 5-ms sine-squared and cosine-squared onset and offset ramps, and a flat spectrum (within 2 dB) within their pass band between 0.5 and 20 kHz, and were digitally generated in MATLAB. The signals were sent to a real-time processor (RP2.1 System3; Tucker-Davis Technologies) at a sampling rate of 48 to



**Figure 2.** Input-output transformation of the robot arm and setup. **A**, Schematization of the experimental set-up. The subject sits erect on a comfortable chair in the center of a circle described by the custom-made light-weight robot arm. The motor driving the robot arm was fixed at the subject’s zenith. The target sound emanated from the loudspeaker, which was mounted at the end of the robotic arm that horizontally rotated the speaker in a pseudo-randomly selected direction (Eq. 2) along a circular trajectory (radius, 1.15 m). **B–D**, Input-output traces [programmed stimulus,  $\hat{\alpha}_n(t)$ , blue, Eq. 2; and motor movement,  $m_n(t)$ , black] and the associated amplitude spectra of the input and output signals. Note that the robot arm added some additional frequency components (red arrows point to examples) to the programmed stimulus.

828.25 Hz. After attenuation by custom-built amplifiers, the audio signal was sent to the loudspeaker, which was moved in a pseudo-random, unpredictable direction (clockwise and anticlockwise). Stimulus coordinates ranged from  $-30^\circ$  to  $+30^\circ$  in azimuth, and at  $0^\circ$  in elevation. All stimuli were well discernible, and kept at a fixed intensity level of 55-dB SPL (A-weighted). Absolute free-field sound levels were measured, with the Brüel & Kjær BK2610 sound amplifier and a Brüel & Kjær BK4144 microphone, at the location of the subject’s head.

The buzzing sounds coming from the activated motor were near 50 dBA at the subject’s ears, and always came from the subject’s zenith,  $90^\circ$  away from the horizontal plane. These sounds did not provide any cue about stimulus location or direction. We tested this qualitatively while the motor was activated, but without playing a target sound. When the target sound played, the motor sounds did not interfere with the listeners’ sound-localization abilities.

The programmed sound-source movements consisted of a linear combination of five sines, digitally generated and stored as a wav-file in MATLAB, to be subsequently sent as a command movement to the robotic arm. Stimulus generation was performed as follows. First, the dynamic sound-source locations,  $\alpha_n(t)$ , for stimulus,  $n$ , with  $n = [1:30]$ , were defined as:

$$\alpha_n(t) = \sum_{k=1}^5 A \cdot \sin(2\pi t f_k + \varphi_{n,k}). \quad (1)$$

The frequency components,  $f_p$ , were fixed multiples of the fundamental frequency,  $f_0 = 0.05$  Hz, i.e.,  $f_p = p \times f_0$ , with  $p = [2,3,7,13,21]$ . The stimuli thus had a period that corresponded to the total trial length of 20 s.

Each component in Equation 1 had a constant amplitude,  $A$ , while its phase,  $\varphi_{n,k}$ , was selected at random between  $[0,2\pi]$ . Because of the latter, the maximum

amplitude of  $\alpha_n(t)$  changed too. We therefore normalized each stimulus by its peak amplitude to a peak excursion of  $30^\circ$ , which resulted in a pseudorandom trial-to-trial variation of the component amplitudes,  $A_k = A/\max(|\alpha_n(t)|)$ , of the harmonics in the stimuli. In this way, the stimulus movements were unpredictable for the subject.

The actual robot movements (i.e., the subjects’ true stimulus motion),  $m_n(t)$ , were measured with a search coil attached to the speaker, and resulted in a slightly nonlinear, and filtered, transformation,  $h(\alpha)$ , of the command input (Eq. 1) to the robotic arm. The true stimulus motion of the motor,  $m_n(t)$ , was thus described by the following:

$$\hat{\alpha}_n(t) = \sum_{k=1}^5 A_k \cdot \sin(2\pi t f_k + \varphi_{n,k}) \text{ and } m_n(t) = h[\hat{\alpha}_n(t)]. \quad (2)$$

Figure 2 shows the robot’s response for one of the stimuli, and the associated frequency components that resulted from the stimulus-to-movement transformation,  $h$ . Note that because of the nonlinear characteristic of  $h(\alpha)$ , the actual motion of the speaker could contain some additional harmonics, e.g., at  $p = [1,5,15]$ , i.e., at 0.05, 0.25, and 0.75 Hz.

### Psychophysics

Subjects performed two psychophysical tasks in different sessions. Both sessions started with a calibration procedure for the head-mounted coil. The first session assessed baseline sound-localization performance in the azimuth and elevation directions, by means of a standard sound-localization task, consisting of 150 trials. In the second session, the subject performed the auditory pursuit paradigm. The latter consisted of thirty different trials of 20 s each. To prevent fatigue that would potentially degrade performance, this session lasted  $\sim 25$  min. The

localization and pursuit tasks were executed under open-loop conditions, in darkness, and without any kind of verbal or visual feedback. For safety reasons, the subject was observed by the experimenter through an infrared camera that was placed in the experimental room.

### Calibration procedure

To obtain the head-position data for the calibration procedure, the subject accurately pointed the head-fixed laser pointer (see above, Experimental setup) toward 56 LED locations distributed over the two-dimensional frontal hemifield. A feedforward three-layer neural network was trained to map the measured endpoints into degrees azimuth and elevation of the LEDs. This neural network was subsequently used for offline calibration of the head-coil signals obtained from the subjects' head-movement responses to the auditory stimuli in the localization task and pursuit experiment.

### Static sound-localization task

To measure the baseline sound-localization performance of the subjects, a standard sound-localization task was performed. The subjects were instructed to point at a straight-ahead fixation LED and push a button whenever they felt ready. After the button press, the fixation LED extinguished, and  $\sim 200$  ms later an auditory Gaussian white noise burst with a duration of 150 ms was presented at a fixed intensity of 55 dBA at a randomly selected location in the subjects' two-dimensional frontal hemifield. The subjects had to point the head, as quickly and as accurately as possible, to the perceived sound location.

### Auditory pursuit task

Subjects were instructed to point at a straight-ahead fixation LED and subsequently pushed a button whenever they felt ready. Immediately after the button press, the fixation LED went off, and  $\sim 200$  ms later the auditory stimulus, consisting of 20 s continuous Gaussian white-noise, appeared at an intensity of 55 dBA. As soon as the stimulus was heard, the subject had to point and track the sound source with the head-mounted-LED (which was continuously on), as accurately as possible.

### Data analysis

All data analysis procedures were performed in MATLAB R2018b (The MathWorks). The coordinates of the moving sound and the head-movement responses were expressed in the double-pole azimuth-elevation coordinate system, in which the origin coincides with the center of the head (Knudsen and Konishi, 1979). The analysis of head movements was performed offline with custom-made software that automatically detected head displacements and saccades in the calibrated data. Detected movements and saccades were checked visually without stimulus information to the experimenter, and onset and offsets could be corrected manually, if needed.

### Sound localization

We quantified the static sound-localization performance of each subject by linear regression on the stimulus-response relations for azimuth and elevation:

$$\alpha_r = b \cdot \alpha_t + a \text{ and } \varepsilon_r = d \cdot \varepsilon_t + c, \quad (3)$$

with  $\alpha_r$ ,  $\alpha_t$ ,  $\varepsilon_r$ , and  $\varepsilon_t$  the response azimuth, target azimuth, response elevation, and target elevation, respectively. Fit parameters,  $a$  and  $c$ , are the response biases (offsets, in degrees), whereas  $b$  and  $d$  are the response gains (slopes, dimensionless) of the azimuth and elevation responses. The parameters  $a, b, c, d$  were found by minimizing the mean-squared error (MSE) of Equation 3 (Press et al., 1992). From each linear fit, we also determined the correlation coefficient between data and fit, the mean absolute error, and the SD of the residuals of the responses.

We verified normal localization performance in azimuth and elevation of all subjects, with gains close to 1.0, biases close to  $0^\circ$ , and high correlation coefficients, typically exceeding 0.9; here, we will not report further on the results of these standard control localization experiments.

### Modelling the pursuit responses

Sound-source pursuit in the horizontal plane was quantified by the frequency content of the stimulus-response waveforms. During pursuit, subjects did not make appreciable vertical head movements. To compare the significant frequency components in the stimuli with those present in the subjects' responses we applied the fast Fourier transform to the stimulus and response signals. From the resulting spectra, we determined the gain-shift and phase-shift characteristics of the responses with respect to the measured stimulus movement for each trial.

We subsequently modelled the pursuit transfer characteristic of each trial,  $n$ , by a second-order linear filter, in series with a delay,  $T_D$ , as has been done before for the ocular visual-pursuit system (Krauzlis and Lisberger, 1994; Barnes, 2008). In the Laplace domain, the transfer characteristic of the APS,  $H_{APS,n}(s)$ , is then given by the following:

$$H_{APS,n}(s) = G_0 \cdot \exp(-T_D \cdot s) \cdot \frac{\omega_{C,n}^2}{\omega_{C,n}^2 + 2\zeta_n \omega_{C,n} s + s^2}, \quad (4)$$

with  $\omega_{C,n} = 2\pi f_C$  the angular resonance frequency of the (undamped) system,  $\omega_{C,n} = 2\pi/T_{C,n}$  (with  $T_{C,n} = 1/f_C$  the system's undamped time constant),  $G_0$  the system's steady-state gain (at  $s=0$ ), and  $\zeta_n$  the system's damping ratio (dimensionless). The delay,  $T_D$ , was determined by brute-force fitting, and clamped at a fixed value, separate for each subject (values ranged between 10 and 98 ms; mean  $\pm$  SD:  $42 \pm 35$  ms). Similarly, we clamped  $G_0 = 1.0$  at 0 Hz. The remaining two parameters of the model that were free to vary across trials, namely the damping ratio and the system's time constant, were found by

MATLAB's *procest* routine (process estimation). The amount of “damping” of the model is usually quantified by its so-called quality factor:

$$Q_n \equiv \frac{1}{2\zeta_n}. \tag{5}$$

The impulse response in the time domain of the model is given by the following:

$$h_{APS,n}(t) = \mathcal{L}^{-1}[\exp(-T_D \cdot s) * \mathcal{L}^{-1} \left[ \frac{\omega_{C,n}^2}{\omega_{C,n}^2 + 2\zeta_n \omega_{C,n} s + s^2} \right]], \tag{6}$$

(with \* signifying convolution in the time domain). Noting that

$$\mathcal{L}^{-1}[\exp(-as)] = \delta(t - a) \text{ and } \mathcal{L}^{-1} \left[ \frac{b}{(s + c)^2 + b^2} \right] = \exp(-ct) \cdot \sin(bt) \cdot u(t),$$

with  $u(t)=1$  for  $t \geq 0$  the Heaviside unit-step function, Equation 6 yields:

$$h_{APS,n}(t) = T_{C,n} \beta_n \cdot \exp(-\omega_{C,n} \cdot \zeta_n (t - T_D)) \cdot \sin(\omega_{C,n} \beta_n \cdot (t - T_D)) \text{ for } \zeta_n < 1, \tag{7a}$$

$$h_{APS,n}(t) = T_{C,n} \gamma_n \cdot [\exp(-\omega_{C,n} (\zeta_n - \gamma_n) (t - T_D)) - \exp(-\omega_{C,n} (\zeta_n + \gamma_n) (t - T_D))] \text{ for } \zeta_n > 1, \tag{7b}$$

where  $\beta_n \equiv \sqrt{1 - \zeta_n^2}$ ,  $\gamma_n \equiv \sqrt{\zeta_n^2 - 1}$  and  $t \geq T_D$ .

To test how well this simple linear model accounts for the response data, we calculated the predicted head-position responses,  $H_{pred}(t)$ , of the model for each subject and trial,  $n$ , by convolving the measured stimulus movement,  $m_n(t)$ , with the fitted impulse response function of Equation 7 by the following:

$$H_{Pred,n}(t) = \int_0^\infty m_n(t - \tau) \cdot h_{APS,n}(\tau) \cdot d\tau \text{ for } n = 1 - 30. \tag{8}$$

The gain characteristic of the underdamped response of the system ( $Q_n > 0.5$ , or  $0 < \zeta_n < 1$ ) reaches its maximum value at frequency:

$$\omega_{max,n} = \omega_{C,n} \cdot \sqrt{1 - \frac{1}{4Q_n^2}} \text{ with } G_{max}(\omega_{max,n}) = \frac{4Q_n^2}{\sqrt{16Q_n^2 - 3}}. \tag{9}$$

The location of the peak approaches  $\omega_{C,n}$  for large  $Q_n$  (with amplitude  $Q_n$ ). In case that  $\omega_{C,n} = \omega_C$  (constant, as is found in our data),  $\omega_{max,n}$  decreases with decreasing  $Q_n$ , together with the system's maximum gain.

At  $Q_n = 0.5$  the maximum gain (1.0) is reached at  $\omega_{(max,n)} = 0$  (critically damped). Thus, the system's

effective time constant increases with decreasing  $Q_n$ , while its overshoots decrease in size.

### Statistics

To quantify and evaluate how well the model represented the subjects' responses, we calculated Pearson's linear correlation coefficient,  $r$ , between the measured and predicted head-position time traces. We also determined the coefficient of determination,  $r^2$ , which quantifies the variability accounted for by the model. In addition, we calculated the MSE of the model for each trial. The effect sizes and confidence intervals (CIs) are reported as effect size (CI width lower bound; upper bound).

We also estimated the mean perceived absolute pursuit error (in degrees) during each trial,  $n$ , for each subject,  $s$ , by calculating:

$$MAE_{n,s} = \frac{1}{20} \int_0^{20} |m_n(t) - H_{n,s}(t + T_{Opt,n,s})| \cdot dt, \tag{10}$$

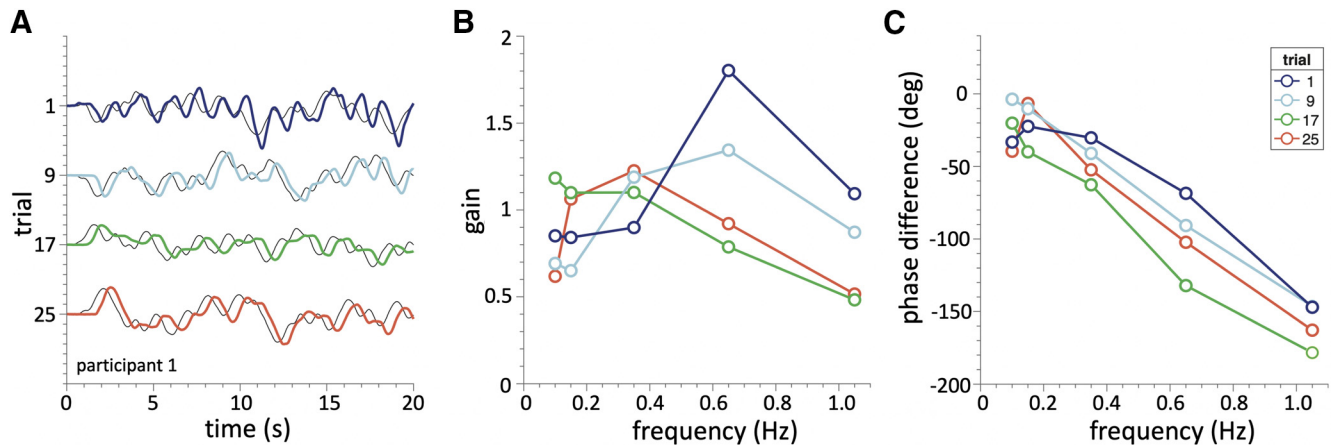
with  $H_{n,s}(t + T_{Opt,n,s})$  the measured head movement of subject  $s$  in trial  $n$ , leading by  $T_{Opt,n,s}$  ms with respect to the stimulus movement,  $m_n(t)$ . This delay was found by a brute-force search for the value that would minimize the mean absolute error (Eq. 10) between target and head movement during that trial (see Figs. 7, 8A,B).

To quantify a potential change in the model parameters across trials, we fitted a hierarchical linear regression model, obtaining slopes and intercepts for each subject and for the group as a whole (Kruschke, 2014) via the sampling program JAGS through MATLAB (Plummer, 2003; Steyvers, 2011). We report the mean and 95% highest-density intervals (HDIs) for the slopes of the fitted lines.

## Results

### Example auditory pursuit

We first illustrate the pursuit behavior of our subjects by showing some representative stimulus-response traces to four unpredictable sound-source movements and their associated transfer characteristics for subject S1 (Fig. 3). A qualitative inspection of the traces (Fig. 3A) indicates that the subject's head movements lagged the stimulus movements during the entire trial, with an average delay of ~300 ms. Leading movements were not observed in these trials. The smooth-pursuit head movements were in the direction of stimulus motion, and corrective fast head saccades were not detected at any stage of the stimulus presentation. To determine the transfer gain and phase characteristics for the five major frequencies in the motor movements (Fig. 3B,C), we applied the fast Fourier transform to the stimulus and response traces. The subject's gains (response amplitude divided by stimulus amplitude at each stimulus frequency) tended to fall off at higher frequencies, with its highest value at intermediate frequencies, suggesting a bandpass response behavior. A qualitative inspection of these gain and phase



**Figure 3.** Examples of auditory pursuit to unpredictable sound movements. Four representative trials from subject S1. **A**, Thin black traces: stimulus movement; colored traces: head movements. The stimulus movement contained five discrete harmonics with random phases (Eqs. 1, 2). **B**, **C**, Transfer characteristics (gain, and phase, in degrees; linear scale) of the stimulus-response relations at the five stimulus frequencies (open dots) for each trial. Both characteristics vary with trial number.

characteristics suggests that they systematically change with trial number.

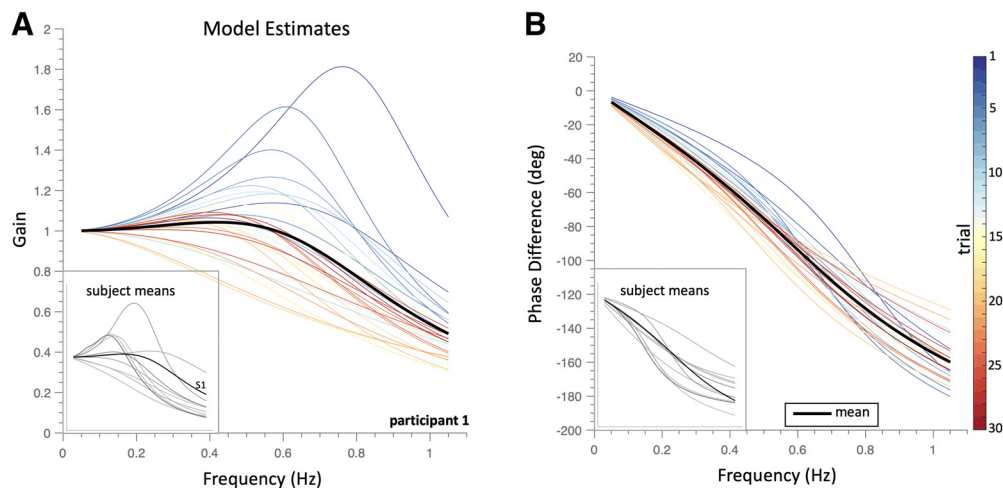
### APS identification

Based on the qualitative observations in Figure 3, we modelled the system's responses in each trial by the second-order filter characteristic of Equation 4. Figure 4 illustrates the individual fitted transfer functions from subject S1 for all 30 trials (thin colored lines), together with the averaged gain (dimensionless; Fig. 4A) and phase characteristics (in degrees; Fig. 4B, bold black lines).

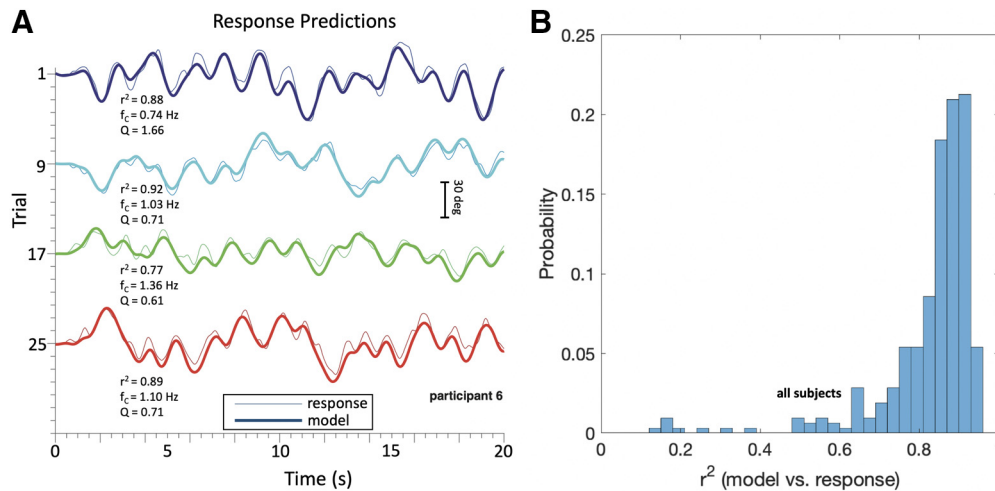
The results suggest a systematic change of the model parameters with trial number: the size and location of the peak of the amplitude characteristic gradually shifted from higher to lower frequencies with trial number (Fig. 4A). Similarly, the phase characteristics changed gradually with trial number from higher to lower frequencies at

each given lag. We obtained similar results for the other participants (Fig. 4, insets).

To test whether the simple feedforward second-order filter was able to account for the full stimulus-response behavior of the participants, we calculated the predicted responses through convolution (Eq. 8) of the model's impulse response function for each trial,  $h_{APS,n}(t)$  (Eq. 7), and the measured stimulus movement,  $m_n(t)$  (Eq. 2). Figure 5A shows four representative examples of the measured (thin colored traces) and predicted responses (bold colored traces) of participant S6. Figure 5B provides the coefficients of determination ( $r^2$ ) for all 11 subjects and trials. These results show that the simple model of Equations 4, 7 provided an excellent description of the response data for all participants and the far majority of trials, with the mode of the distribution at  $r^2 = 0.88$  (i.e., 88% of variance explained; across subjects: mean  $\pm$  SD:  $0.83 \pm 0.10$ ).



**Figure 4.** Estimated model fits for participant S1 (Eq. 4). Gain-characteristics (**A**) and phase-characteristics (**B**) for the stimulus-response data of all 30 trials as function of frequency. The thin curves show the results for each trial, color coded by trial number. The black solid lines correspond to the average transfer characteristic across trials. Note the systematic change of the gain and phase characteristics with trial number. Insets, average gain and phase characteristics for all 11 subjects.



**Figure 5. A**, Measured (thin traces) and predicted (thick traces; Eq. 8) head-movement responses for four representative trials of subject S6. Each trace shows head orientation as a function of time. Optimal fit parameters ( $f_C$ ,  $Q$ , and goodness of fit,  $r^2$ ) are provided for each trial. Note the good correspondence between measured and predicted movements. The subject's response delay was clamped at  $T_D = 98$  ms for all trials. **B**, Histogram of the coefficients of determination,  $r^2$ , for all 315 trials (11 subjects). The mode of the distribution lies at  $r^2 = 0.88$ , which indicates an excellent fit of the data (variance explained).

### Adaptive changes in auditory pursuit

Figure 4 suggests a systematic change of the model characteristic as a function of trial number. To quantify this trend in the pursuit behavior of all subjects, we performed a linear regression analysis (for details, see Materials and Methods) on the two free parameters of the model: its center frequency,  $f_{C,n}$ , and the quality factor,  $Q_n$ , as a function of trial number,  $n$  (Fig. 6).

We first illustrate the results for subject S3 in Figure 6A, C. The center frequency (Fig. 6A, blue) did not change systematically during the experiment, with the data (open circles) scattering around the average value of  $f_C \approx 0.6$  Hz. The optimal regression line through the data (solid line) had a slope close to zero (slope =  $-0.8 \times 10^{-3}$ ; 95% HDI =  $[-4.4, +2.8] \times 10^{-3}$ ). The  $f_C$  values predicted by the linear model at the start of the experiment (at trial 1, inset histogram and 95% HDI bar on the left) and at the end of the experiment (trial 30, inset histogram and bar on the right) were also similar (the probability densities indicated by the histograms and bars for trial 1 and 30 overlapped considerably).

This subject's quality factor (Fig. 6C, open squares) seemed to decrease on average although variability between individual trials was considerable. Nevertheless, the optimal slope was clearly non-zero ( $-13 \times 10^{-3}$ ; 95% HDI =  $[-21, -7] \times 10^{-3}$ ) and the variability in the fitted lines was low (as reflected by the shaded area indicating the 95% HDI of the regression). Similarly, the most likely  $Q$ -values at the first and last trials did not overlap at all (compare left and right histograms and bars).

Very similar results held for all 11 subjects (Fig. 6B,D). The center-frequency data averaged across subjects (Fig. 6B, open circles) scattered around a value of  $f_C \approx 0.6$  Hz, and did not vary significantly during the experiment. The slope of the optimal average regression line was near-zero ( $-0.5 \times 10^{-3}$ ; 95% HDI =  $[-3.5, +2.8] \times 10^{-3}$ ). In contrast, the average quality factor (Fig. 6D, open squares)

decreased substantially as the experiment progressed, from values higher than 1.0 in the early phase of the experiment (indicating an under-damped, bandpass behavior) toward the end of the session. This change was also reflected in the group regression slope of  $-10 \times 10^{-3}$  (95% HDI =  $[-15, -5.5] \times 10^{-3}$ ). Overall, the predicted total change of the quality factor across the 30 trials was  $-0.30$ , a nearly 27% difference.

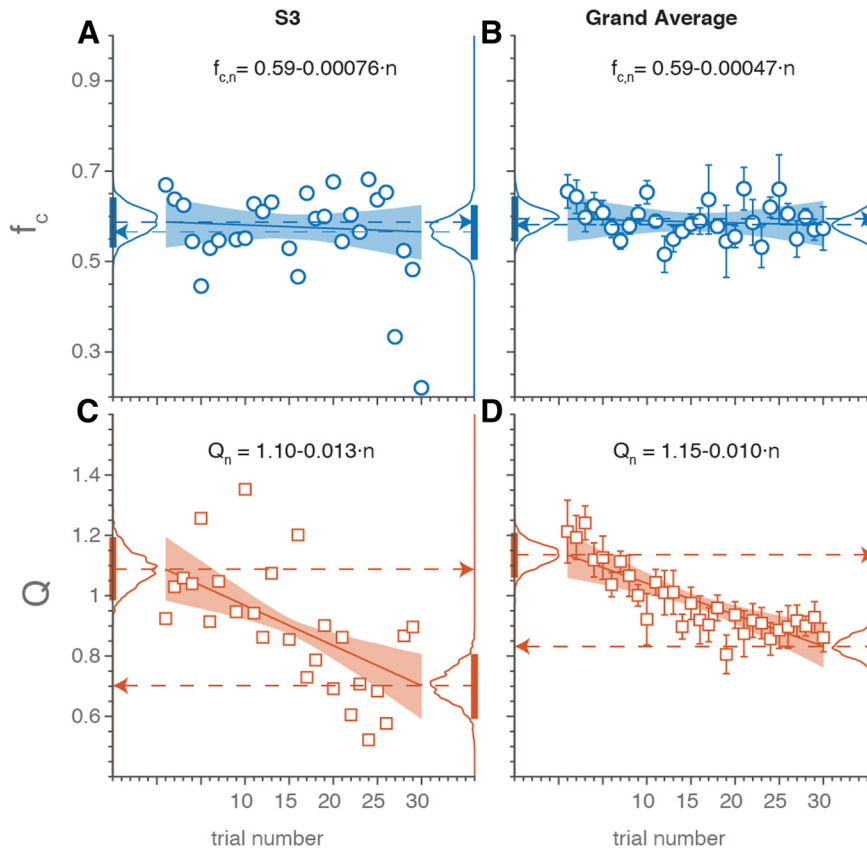
To test whether the change in the response characteristics would lead to improved pursuit performance we calculated the mean absolute localization error (in degrees) for each trial (Eq. 10). Figure 7 shows the results of this analysis for each participant (color coded) and for the average behavior across subjects (black dots). The data show that the mean absolute error across participants remained constant at  $4.8^\circ$  (SD  $1.7^\circ$ ) throughout the experiment.

### Position error versus velocity error

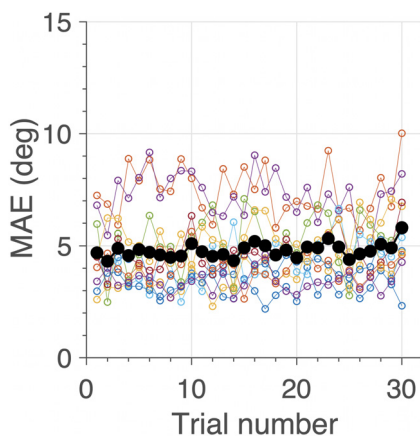
The head-movement responses of our subjects contained the same spectral motion components as the stimulus, which suggests that the responses may have been driven by sound-source velocity. However, Rashbass (1961) noted for visual pursuit that there is a theoretical possibility that the pursuit system samples target positions at a sufficiently high rate that exceeds the spectral bandwidth of the response system. In that case, discrete position sampling is indistinguishable from smooth pursuit of target velocity. To rule out the former would require a paradigm in which a directional change in position error is dissociated from the direction of the smooth target movement. To our knowledge, such an experiment has not yet been conducted for auditory pursuit.

However, the random stimulus movements in the current experiment might in principle allow for some dissociation between these two variables. To check for this





**Figure 6.** Adaptation of auditory pursuit. **A, B**, The model's resonance frequency,  $f_c$  and quality factor,  $Q$  (**C, D**), as a function of trial number for subject S3 (**A, C**), and pooled across all subjects (**B, D**). Shaded areas indicate the 95% HDI for the regression lines through the data. For the pooled results, a hierarchical regression analysis was performed, which accounts for the individual differences. Histograms on the left and right of each panel show the distributions of the respective parameter at the start (trial 1), and end of the experiment (trial 30), respectively, with their HDI (bars). The system's resonance frequency did not change systematically with trial number, while the quality factor decreased systematically, and highly significantly, throughout the experiment. The equations in each panel denote the optimal regression results.



**Figure 7.** MAE (in degrees; Eq. 10) as function of trial number for each subject (different colors), and the grand averages across subjects (solid black dots). There was no trend for a change (neither positive, nor negative) in the MAE with trial number.

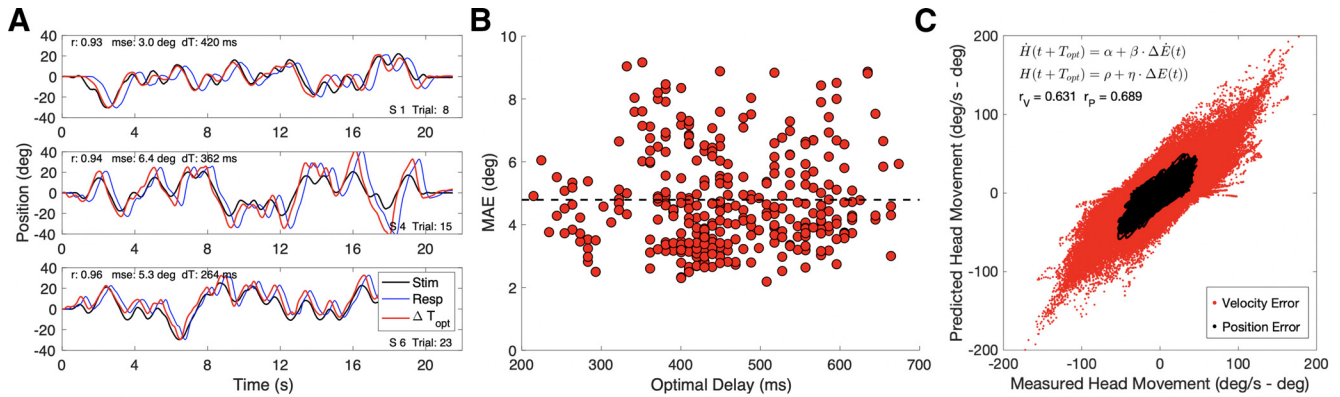
possibility, we performed a trial-by-trial regression analysis on the instantaneous head velocity versus head-velocity error, and on the current head position versus head-position error, respectively:

$$\dot{H}_{n,s}(t + T_{Opt,n,s}) = \alpha + \beta \cdot \Delta \dot{E}_{n,s}(t), \quad (11a)$$

$$H_{n,s}(t + T_{Opt,n,s}) = \varrho + \eta \cdot \Delta E_{n,s}(t), \quad (11b)$$

where position error is defined as  $\Delta E_{n,s}(t) = m_n(t) - H_{n,s}(t + T_{Opt,n,s})$  and head-velocity error by  $\Delta \dot{E}_{n,s}(t) = \dot{m}_n(t) - \dot{H}_{n,s}(t + T_{Opt,n,s})$ . Here,  $\alpha$ ,  $\beta$ ,  $\rho$ , and  $\eta$  are the regression parameters obtained for the entire data set, and  $m_n(t)$  is the sound-movement trajectory of trial  $n$ .  $T_{Opt,n,s}$  is the optimal delay of the head movement found in trial  $n$  for subject  $s$ .

Figure 8 presents the results of this analysis. Figure 8A, B illustrates the procedure of finding the optimal time-shift of the head trajectory,  $T_{Opt,n,s}$ , such that it aligned best with the stimulus trajectory during the trial [yielding the smallest mean absolute pursuit error (MAE)], for three different subjects and trials. Note the very high correlations between stimulus movement and time-shifted head



**Figure 8.** **A**, Three example trials for three different subjects, showing the stimulus movement (black), the head movement (blue), and the time-shifted head movement that aligns best with the stimulus (red).  $r$ : correlations between stimulus and time-shifted head-movement traces are very high. MSE: the mean-absolute errors in degrees. dT: the applied optimal shift (in milliseconds). **B**, Distribution of the mean absolute errors and optimal delays for all trials and subjects ( $N=330$ ). **C**, Measured instantaneous (optimally shifted) head velocity versus the predicted head velocity (Eq. 11b; red) and measured head position versus predicted head position (Eq. 11a; black). Both models correlate well ( $r_V = 0.63$  vs  $r_P = 0.69$ ; total number of data points:  $N = 726,000$ ; see also Table 1). The errors themselves are uncorrelated ( $r = -0.01$ ).

movement ( $r > 0.9$ , for 2200 data points). Figure 8B shows the joint distribution for all trials and subjects of best delays and minimum MAEs (the latter also shown in Fig. 7A). Figure 8C shows the predictions of Equations 11a (red) and 11b (black) for the entire dataset (11 subjects  $\times$  30 trials  $\times$  2200 samples = 726,000 points). Both regressions yield a high correlation, indicating that subjects followed the pseudo-random stimulus trajectories quite faithfully (Table 1). Note the slightly higher correlation for the position error predictor ( $r_P = 0.69$ ) than for the velocity error predictor ( $r_V = 0.63$ ), while the two errors themselves were uncorrelated ( $r = -0.01$ ). Moreover, the head velocity was unrelated to the instantaneous sound position ( $r < 0.001$ ; see Table 1). Thus, the head-velocity error and head-position error both described the head-movement data (velocity and position) about equally well.

## Discussion

Our results show that subjects smoothly track broadband sounds, moving along unpredictable trajectories in the horizontal plane, with remarkable accuracy. The frequency spectrum of their head-movements contained the same dominant frequencies as the source-movements.

**Table 1: Linear regression on instantaneous head position (H) as function of the head-position error ( $\Delta E$ ; Eq. 11b), and on head velocity ( $\dot{H}$ ) as function of the velocity error ( $\Delta \dot{E}$ ; Eq. 11a)**

Regression	Offset	Slope	Correlation
$H = \alpha + \beta \cdot \Delta E$	-0.98	-1.13	0.55
$\dot{H} = \rho + \eta \cdot \Delta \dot{E}$	0.00	-0.96	0.63
$H = a + b \cdot \Delta \dot{E}$	0.16	-0.009	0.01
$\dot{H} = c + d \cdot \Delta E$	0.20	0.15	0.03

Both regressions yield high gains and correlations. Head position is unrelated to the velocity error, and, conversely, head velocity is independent of the position error. All regressions were performed on a subset of 48,400 data points (trajectories were sampled in steps of 15 samples, because of memory limitations).

We described the tracking responses by a simple feedforward second-order linear filter and found that its damping gradually increased during the experimental session. A slightly underdamped bandpass response in early trials, turned into an overdamped response with a longer effective time constant (Eq. 9) and lag (Figs. 4, 6). We argue that this behavior reflects an adaptation of the APS to a learned stimulus characteristic. In what follows, we discuss these features in more detail.

## Pursuit accuracy

Although ocular pursuit of visual targets is well documented and modelled (Robinson et al., 1986; Krauzlis and Lisberger, 1994; Barnes, 2008), much less is known about the cranial pursuit of moving sounds. Earlier eye-movement studies have reported poor auditory pursuit performance (gain  $< 0.2$ ; Boucher et al., 2004; Berryhill et al., 2006) and concluded that the auditory system has no specialized motion detectors, because of its coarse spatial acuity compared with vision. However, we argue that lack of accurate eye-movement tracking to moving sounds is not evidence for absence of motion-sensitive auditory processing.

The ocular pursuit system successfully tracks visual targets to reduce retinal slip-velocity through continuous visual feedback (Robinson et al., 1986; Krauzlis and Lisberger, 1994; Lisberger, 2010). While visual-cortical and subcortical areas contain pursuit-sensitive and target-velocity sensitive neurons (Mikami et al., 1986; Dürsteler et al., 1987; Ilg and Thier, 2008), evidence for auditory motion-sensitive cells has been obtained for sensory responses under anaesthetized conditions only (see Introduction). Importantly, however, head-fixed ocular pursuit of sounds does not reduce any error-signal, as the head-centered acoustic information does not change with eye movements. Therefore, it is questionable whether ocular pursuit of sounds may serve as a valid measure for sound-velocity processing. Instead, cranial pursuit does

affect the sound's acoustic cues, in a way that is directly related to self-initiated head movements (Fig. 1).

We here demonstrated that subjects successfully tracked unpredictably moving sounds with smooth head-movements that were always in the stimulus direction at a fixed, idiosyncratic delay (Figs. 3, 8A). Subjects could not anticipate the pseudorandom changes of the target's movement direction as indicated by a constant mean absolute error across trials (Eq. 10) of  $\sim 4.8^\circ$  (Fig. 7). This indicates that the system did not attempt to reduce the perceived pursuit error, presumably because it could not rely on any predictions for these random trajectories. As a result, the head-movement delay during a trial amounted to several hundreds of milliseconds (Fig. 8A,B), which did not change systematically across trials, and was unrelated to the mean absolute error in a trial (Fig. 8B).

Sound-localization experiments with eye-head movements have demonstrated that the auditory system continuously uses eye-movement and head-movement information to update the location of brief sounds (Vliegen et al., 2004; Genzel et al., 2018). We have hypothesized that the APS aims to keep its AF close to the target, just as in visual pursuit (Fig. 1). A putative AF would, by definition, correspond to the region of highest spatial acoustic resolution. For ITDs and ILDs, which vary sinusoidally with the azimuth angle (Blauert, 1997; Van Opstal, 2016), the AF would be around straight-ahead, with  $1.0\text{--}1.5^\circ$  acuity (Mills, 1958). Interestingly, an acoustic spatial fovea is not anatomically represented in the cochlea. It is therefore an abstract, functionally defined concept, neurally generated from binaural integration of different acoustic processing streams. Although it is not yet known how the brain represents an AF, or what the relative contributions of the ITD and ILD pathways are, our data may support its functional existence (see also below).

### Adaptive response behavior

The adaptation of pursuit-responses across trials gradually increased the system's damping. The other parameters of our model (time constant,  $T_c$ , and processing delay,  $T_D$ ) did not change systematically with trial number (Fig. 6). It is not immediately obvious which cost the pursuit system aimed to optimize, as multiple factors may underlie the cost evaluation: position and velocity errors (response accuracy), movement effort (energy consumption), response duration to match target velocity (discount of reward), trajectory smoothness, etc. As experiments were performed under open-loop conditions, participants never obtained exogenous feedback about the true target trajectory, and had to rely entirely on ongoing endogenous processing of acoustic information, together with self-initiated head-movements, and associated vestibular and efference-copy signals. This situation differs radically from classical visual pursuit.

Target-movement trajectories were unpredictable from trial to trial, and also within a trial. Thus, a possible pursuit strategy could have been to generate head-movements through a fixed input-output characteristic. Our data show that in the first couple of trials this characteristic could be well described by a slightly underdamped

impulse response (Eq. 7), for which the frequency-characteristic has maximum gain ( $>1.0$ ) around a cutoff frequency of  $0.6\text{--}0.8$  Hz. Interestingly, visual pursuit to an unexpected change of target velocity is characterized by a similar "ringing" of pursuit eye-velocity at  $3\text{--}3.5$  Hz (Robinson et al., 1986; Krauzlis and Lisberger, 1994).

Remarkably, the APS seemed to extract implicit spectral information from evoked movement trajectories in the stimulus ensemble, and gradually changed its response behavior, such that the underdamped characteristic became near-critically damped, with  $Q \sim 0.75$  ( $Q = 0.5$  is the critically-damped response). To verify that this is indeed the case, future experiments could manipulate the amount of consistent spectral information in the movement trajectories, and test whether it affects the long-term pursuit behavior.

### Response strategy

We propose that for the pseudo-random stimulus set this adaptive response strategy may have optimized a cost that included the system's response-duration and total response effort. Simple estimates of these costs can be made from the model's response characteristic. Figure 9A illustrates the step responses, and Figure 9B the associated impulse responses, of the second-order model of Equation 7, for which we took a fixed resonance frequency of  $f_c = 0.6$  Hz, and the quality factors varying between  $Q = 1.5$  and  $Q = 0.6$  in steps of  $-0.1$ , as obtained in our experiments (Fig. 6).

The duration of the step-response is clearly shortest for the lowest Q factor ( $D_{\min} \sim 1.8$  s). If we assume that the total (rotational kinetic) energy consumption during the head movement is proportional to its absolute squared (angular) velocity, or to the mean spectral power of the system's amplitude characteristic, per unit of angular momentum:

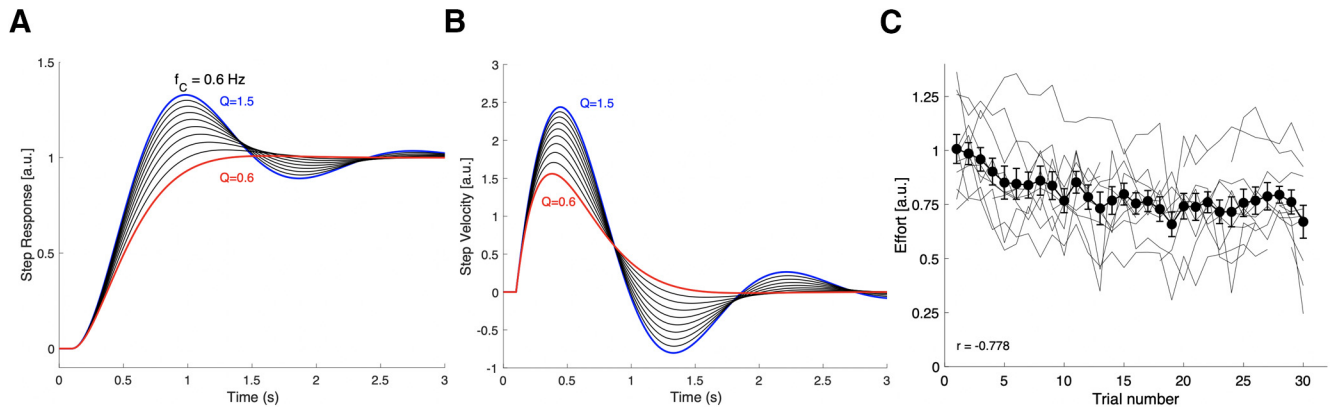
$$E_Q(D) = \left| \int_0^D v^2(t) dt \right| \quad (12a)$$

or

$$\bar{E}_Q = \frac{1}{(\omega_{\max} - \omega_{\min})} \cdot \int_{\omega_{\min}}^{\omega_{\max}} |G(\omega)| d\omega, \quad (12b)$$

then also  $E_Q$  reaches a clear minimum for  $Q = 0.6$  (Fig. 9B). This also holds when the integration window is kept fixed at the minimum  $D = 1.8$  s for all Q values, e.g., for  $Q = 1.5$ ,  $E_{1.5}(1.8) = 2.79$ , and for  $Q = 0.6$ :  $E_{0.6}(1.8) = 1.13$ , which is a reduction of 60%.

In Figure 9C, we plotted the estimated mean absolute spectral power (in arbitrary units) of the fitted gain characteristic over  $0.05\text{--}1.05$  Hz for each trial of all 11 subjects, as a function of trial number. The results show that the total effort estimate indeed decreased systematically during the course of the experiment, by  $\sim 34\%$  (difference



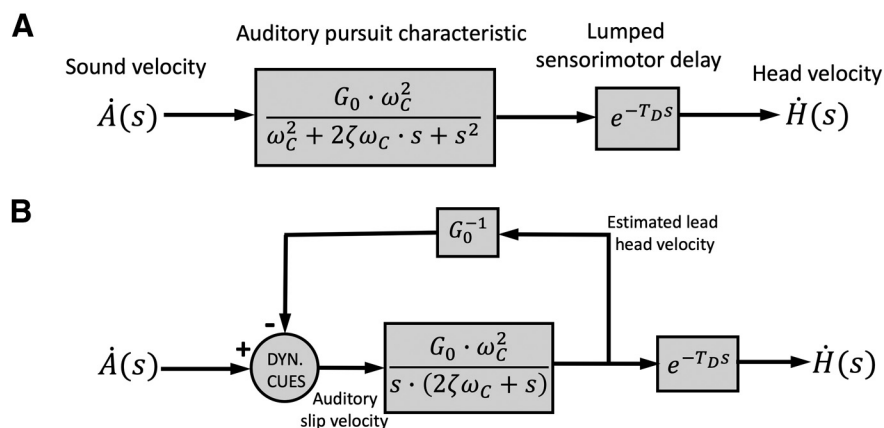
**Figure 9.** **A**, The model’s unit-step responses for 10 values of  $Q$  between  $Q = 1.5$  (blue) and  $0.6$  (red) in decreasing steps of  $-0.1$  (black). **B**, The system’s response velocity (i.e., its impulse response) for the unit-step input to the same  $Q$  values. At  $Q = 0.6$ , the system is near-critically damped and reaches its equilibrium value much faster than for higher  $Q$  values. In addition, the total consumed energy by the system is considerably less at the lower  $Q$ : from  $Q = 1.5$  to  $Q = 0.6$  the total cumulative energy reduction is 60%. **C**, Calculated effort from the fitted gain characteristics of Equation 4, taken as the mean spectral power over 0.05–1.05 Hz for all 11 subjects (thin gray lines), and the mean across subjects (black solid line), plotted as function of trial number. The correlation for the mean is  $r = -0.778$ ; it decreases by 34% between the first and the last trial.

between the first and last trial), with an overall correlation of  $r = -0.78$ .

Minimization of an overall performance cost has also been suggested by others to underlie oculomotor behavior (for eye saccades, Harris and Wolpert, 2006; Sadaghat-Nejad et al., 2019). As the human fovea has a high resolution within only  $1^\circ$  of visual angle, and considerable uncertainty in the retinal periphery, theoretical studies have indicated that the saccadic system aims to optimize speed-accuracy trade-off, to minimize saccade duration at the smallest mean-absolute localization errors (Harris and Wolpert, 2006).

Similar optimization principles appear to hold for human sound localization (Ege et al., 2019). Note that despite the availability of acoustic cues for azimuth and

elevation, veridical localization of a sound is not possible with these cues alone, as the sensory spectrum results from a convolution of source spectrum and pinna cues, both of which are a priori unknown. Thus, the brain cannot be sure about the veridical source direction without making prior assumptions (Middlebrooks and Green, 1991; Van Opstal, 2016). Experiments have suggested that the auditory system uses several priors, learned through experience: for example, (1) each pinna filter refers to a unique elevation angle, and (2) natural source spectra do not resemble the pinna spectra (Hofman and Van Opstal, 1998), (3) not all spectral bands of the pinna filters are equally informative (Zonooz et al., 2019), and (4) not all source locations are equally likely (Ege et al., 2018, 2019). We recently demonstrated that the auditory system



**Figure 10.** Models. Two mathematically equivalent schemes for the APS, on the basis of our results. Both models are represented in the Laplace domain. **A**, Feedforward implementation of Equation 4: a second-order low-pass filter in series with a lumped sensory-motor delay. **B**, Equivalent feedback model, in line with the proposal in Figure 1; the feedback path carries an internal estimate of the output head velocity with a scalar gain,  $1/G_0$ , and lead,  $T_D$ . The feedforward path has a pole in zero (pure integrator), and at  $s = -2\zeta\omega_C$  (i.e., a leaky integrator with time constant,  $T_{FF} = (2\zeta\omega_C)^{-1}$ ). The feedback comparator computes the auditory slip velocity, which in this model would be given by  $v_{SLIP,C}(t) = \dot{A}(t) - G_0^{-1} \cdot H(t+T_D)$ . For simplicity, the head-position error is not included in this scheme.

reweighs its spectral and source-location priors within the same experimental session, without exogenous visual feedback (Zonooz et al., 2018; Ege et al., 2019), suggesting that the brain combines the acoustic input across trials, in combination with its own head-orienting commands, to update its priors. Our results provide evidence for a similar strategy when tracking moving sounds.

### Neural implications

The success of the simple linear feedforward model in predicting auditory pursuit in azimuth does not exclude the possibility that the system may actually be driven by dynamic feedback, in which the neural estimates of craniocentric source velocity and source position result from the combined effect of the true target velocity and position with the self-generated signals related to head-velocity and change in head position, like suggested in Figure 1. A pursuit system would aim to minimize the estimated auditory foveal slip-error, by ensuring that the instantaneous target estimate remains close to the representation of an AF. For that, it should ensure that both the head velocity should be similar to target velocity, and that the AF should be close to the target position. The analysis shown in Figure 8C suggests that the craniocentric velocity and position errors both contribute strongly to the head-pursuit behavior.

A putative example of a feedback system, incorporating our results, is illustrated in Figure 10B.

Note that in visual pursuit, feedback is automatically implemented, as retinal slip is locked to the moving eye. This is less trivial for auditory pursuit, as an auditory (spatial) fovea is not linked to the basilar membrane; its representation should result from neuro-computational mechanisms. Figure 1 indicates the major neural pathways and computational stages for tracking moving sounds in azimuth. Moving sounds produce dynamic changes in the high-frequency ILDs and low-frequency ITDs, processed in binaural brainstem pathways that terminate in the Superior Olivary complex (LSO and MSO; Yin, 2002). Together, the outputs of these pathways converge on IC, a central hub for spatial and spectral-temporal processing of sounds (Groh et al., 2001; Casseday et al., 2002; Zwiers et al., 2004; Versnel et al., 2009). The IC would therefore be the prime target to study tuning to head-centered sound-source velocity and position error, for which some evidence has been obtained from dichotic experiments in anesthetized cats (Al'tman et al., 1985; Bekhterev, 2003), owls (Wagner and Takahashi, 1992), guinea pigs (Ingham et al., 2001), and bats (Pollack, 2012).

Further evidence from cats (Toronchuk et al., 1992), rats (Doan and Saunders, 2003), and humans (Kreitewolf et al., 2011) suggests that also auditory-cortical areas may be responsive to sound velocity. We conjecture that our results may hint at the interesting possibility that these cortical cells could instead encode auditory slip-error in velocity and position with respect to the AF (Figs. 1, 10B), in a similar way as ocular pursuit-responses in visual-cortical areas (Mikami et al., 1986; Dürsteler et al., 1987). To demonstrate this, however, will require

electrophysiological recordings from behaving animals, trained to track moving sounds with the head, which have so far not been performed.

Indeed, as single-unit recordings have demonstrated clear behavioral correlates at different stages in the monkey auditory system (Groh et al., 2001; Zwiers et al., 2004; Massoudi et al., 2013, 2014), we propose that inclusion of the full action-perception cycle is essential to understand the neural processing of moving sounds (Van Opstal, 2016).

### References

- Al'tman YA, Kudryatseva II, Radionova EA (1985) The pattern of response of the inferior colliculus of the cat during the movement of a sound source. *Neurosci Behav Neurophysiol* 15:318–324.
- Barnes GR (2008) Cognitive processes involved in smooth pursuit eye movements. *Brain Cogn* 68:309–326.
- Bekhterev NN (2003) Evoked potentials of the cat inferior colliculus to acoustic stimuli simulating sound source movement with different velocities in opposite directions (in Russian). *Russ Fiziol Zh Im M Sechenova* 89:657–666.
- Beitel RE (1999) Acoustic pursuit of invisible moving targets by cats. *J Acoust Soc Am* 105:3449–3453.
- Berryhill ME, Chiu T, Hughes HC (2006) Smooth pursuit of nonvisual motion. *J Neurophysiol* 96:461–465.
- Blauert J (1997) *Spatial Hearing. The psychophysics of human sound localization*, Ed 2. Cambridge: The MIT Press.
- Boucher L, Lee A, Cohen YE, Hughes HC (2004) Ocular tracking as a measure of auditory motion perception. *J Physiol Paris* 98:235–248.
- Brimijoin WO, Boyd AW, Akeroyd MA (2013) The contribution of head movement to the externalization and internalization of sounds. *PLoS One* 8:e83068.
- Carlile S, Leung J (2016) The perception of auditory motion. *Trends Hear* 20:2331216516644254.
- Casseday JH, Fremouw T, Covey E (2002) The inferior colliculus: a hub for the central auditory system. In: *Integrative functions in the mammalian auditory pathway* (Oertel D, Fay RR, Popper AN, eds), pp 238–318. New York: Springer.
- Crum PAC, Hafter ER (2008) Predicting the path of a changing sound: velocity tracking and auditory continuity. *J Acoust Soc Am* 124:1116–1129.
- Doan DE, Saunders JC (2003) Sensitivity to simulated directional sound motion in the rat primary auditory cortex. *J Neurophysiol* 81:2075–2087.
- Dürsteler MR, Wurtz RH, Newsome WT (1987) Directional pursuit deficits following lesions of the foveal representation within the superior temporal sulcus of the macaque monkey. *J Neurophysiol* 57:1262–1287.
- Ege R, Van Opstal AJ, Van Wanrooij MM (2018) Accuracy-precision trade-off in human sound localisation. *Sci Rep* 8:16399.
- Ege R, Van Opstal AJ, Van Wanrooij MM (2019) Perceived target range shapes human sound localization behavior. *eNeuro* 6: ENEURO.0111-18.2019.
- Firzlaff U, Schuller G (2001) Cortical representation of acoustic motion in the rufous horseshoe bat, *Rhinolophus rouxi*. *Eur J Neurosci* 13:1209–1220.
- Genzel D, Schutte M, Brimijoin BO, MacNeilage PR, Wiegrebe L (2018) Psychophysical evidence for auditory motion parallax. *Proc Natl Acad Sci USA* 115:4264–4269.
- Grantham DW (1986) Detection and discrimination of simulated motion of auditory targets in the horizontal plane. *J Acoust Soc Am* 79:1939–1949.
- Groh JM, Trause AS, Underhill AM, Clark KR, Inati S (2001) Eye position influences auditory responses in primate inferior colliculus. *Neuron* 29:509–518.

- Harris CM, Wolpert DM (2006) The main sequence of saccades optimizes speed-accuracy trade-off. *Biol Cybern* 95:21–29.
- Harris JD, Sergeant RL (1971) Monaural-binaural minimum audible angles for a moving sound source. *J Speech Hear Res* 14:618–629.
- Hofman PM, Van Opstal AJ (1998) Spectro-temporal factors in two-dimensional human sound localization. *J Acoust Soc Am* 103:2634–2648.
- Hofman PM, Van Riswick JG, Van Opstal AJ (1998) Relearning sound localization with new ears. *Nat Neurosci* 1:417–421.
- Ilg UJ, Thier P (2008) The neural basis of smooth pursuit eye movements in the rhesus monkey brain. *Brain Cogn* 68:229–240.
- Ingham NJ, Hart HC, McAlpine D (2001) Spatial receptive fields of inferior colliculus neurons to auditory apparent motion in free field. *J Neurophysiol* 85:23–33.
- Knudsen EI, Konishi M (1979) Mechanisms of sound localization in the barn owl (*Tyto alba*). *J Comp Physiol A Neuroethol Sens Neural Behav Physiol* 133:13–21.
- Krauzlis RJ (2004) Recasting the smooth pursuit eye movement system. *J Neurophysiol* 91:591–603.
- Krauzlis RJ, Lisberger SG (1991) Visual motion commands for pursuit eye movements in the cerebellum. *Science* 253:568–571.
- Krauzlis RJ, Lisberger SG (1994) A model of visually-guided smooth pursuit eye movements based on behavioral observations. *J Comput Neurosci* 1:265–283.
- Kreitewolf J, Lewald J, Getzmann S (2011) Effect of attention on cortical processing of sound motion: an EEG study. *Neuroimage* 54:2340–2349.
- Kruschke J (2014) *Doing Bayesian data analysis*, Ed 2. San Diego: Elsevier.
- Lisberger SG (2010) Visual guidance of smooth-pursuit eye movements: sensation, action, and what happens in between. *Neuron* 66:477–491.
- Makous JC, Middlebrooks JC (1990) Two-dimensional sound localization by human listeners. *J Acoust Soc Am* 87:2188–2200.
- Massoudi R, Van Wanrooij MM, Van Wetter SMCI, Versnel H, Van Opstal AJ (2013) Stable bottom-up processing during dynamic top-down modulations in monkey auditory cortex. *Eur J Neurosci* 37:1830–1842.
- Massoudi R, Van Wanrooij MM, Van Wetter SMCI, Versnel H, Van Opstal AJ (2014) Task-related preparatory modulations multiply with acoustic processing in monkey auditory cortex. *Eur J Neurosci* 39:1538–1550.
- Middlebrooks JC (1992) Narrow-band sound localization related to external ear acoustics. *J Acoust Soc Am* 92:2607–2624.
- Middlebrooks JC (2015) Sound localization. *Handb Clin Neurol* 129:99–116.
- Middlebrooks JC, Green DM (1991) Sound localization by human listeners. *Annu Rev Psychol* 42:135–159.
- Mikami A, Newsome WT, Wurtz RH (1986) Motion selectivity in macaque visual cortex. I. Mechanisms of direction and speed selectivity in extrastriate area MT. *J Neurophysiol* 55:1308–1327.
- Mills AW (1958) On the minimum audible angle. *J Acoust Soc Am* 30:237–246.
- Musicant AD, Butler RA (1984) The influence of pinnae-based spectral cues on sound localization. *J Acoust Soc Am* 75:1195–1200.
- Oldfield SR, Parker SP (1984) Acuity of sound localisation: a topography of auditory space. II. Pinna cues absent. *Perception* 13:601–617.
- Olsen JF, Suga N (1991) Combination-sensitive neurons in the medial geniculate body of the mustached bat: encoding of relative velocity information. *J Neurophysiol* 65:1254–1274.
- Pallus A, Freedman GE (2016) Target position relative to the head is essential for predicting head movement during head-free gaze pursuit. *Exp Brain Res* 234:2107–2121.
- Plummer M (2003) JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. *Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003)*, Vienna, Austria.
- Poirier P, Jiang J, Lepore F, Guillemot JP (1997) Positional, directional and speed selectivities in the primary auditory cortex of the cat. *Hear Res* 113:1–13.
- Pollack G (2012) Circuits for processing dynamic interaural intensity disparities in the inferior colliculus. *Hear Res* 288:47–57.
- Press WH, Teukolsky AA, Vetterling WT, Flannery BP (1992) *Numerical recipes in C: the art of scientific computing*, Ed 2. Cambridge: Cambridge University Press.
- Rashbass C (1961) The relationship between saccadic and smooth tracking eye movements. *J Physiol* 159:326–338.
- Robinson DA (1963) A method of measuring eye movement using a scleral search coil in a magnetic field. *IEEE Trans Biomed Eng* 10:137–145.
- Robinson DA (1965) The mechanics of human smooth pursuit eye movement. *J Physiol* 180:569–591.
- Robinson DA, Gordon JL, Gordon SE (1986) A model of the smooth pursuit eye movement system. *Biol Cybernet* 55:43–57.
- Sadaghat-Nejad E, Herzfeld DJ, Shadmehr R (2019) Reward prediction error modulates saccade vigor. *J Neurosci* 39:5010–5017.
- Steyvers M (2011) MATJAGS 13 A Matlab interface for JAGS. Available at [http://psiexp.ss.uci.edu/research/programs\\_data/jags/](http://psiexp.ss.uci.edu/research/programs_data/jags/).
- Stumpf E, Torunchuk JM, Cynader MS (1992) Neurons in cat primary auditory cortex sensitive to correlates of auditory motion in three-dimensional space. *Exp Brain Res* 88:158–168.
- Torunchuk JM, Stumpf E, Cynader MS (1992) Auditory cortex neurons sensitive to correlates of auditory motion: underlying mechanisms. *Exp Brain Res* 88:169–180.
- Van Bentum GC, Van Opstal AJ, Van Aartrijk CMM, Van Wanrooij MM (2017) Level-weighted averaging in elevation to synchronous amplitude-modulated sounds. *J Acoust Soc Am* 142:3094–3103.
- Van Opstal AJ (2016) *The auditory system and human sound-localization behavior*, pp 436. Amsterdam: Academic Press.
- Van Wanrooij MM, Van Opstal AJ (2005) Relearning sound localization with a new ear. *J Neurosci* 25:5413–5424.
- Versnel H, Zwiers MP, Van Opstal AJ (2009) Spectrotemporal response properties of inferior colliculus neurons in alert monkey. *J Neurosci* 29:9725–9739.
- Vliegen J, Van Grootel TJ, Van Opstal AJ (2004) Dynamic sound localization during rapid eye-head gaze shifts. *J Neurosci* 24:9291–9302.
- Wagner H, Takahashi T (1992) Influence of temporal cues on acoustic motion-direction sensitivity of auditory neurons in the owl. *J Neurophysiol* 68:2063–2076.
- Wightman FL, Kistler DJ (1989) Headphone simulation of free-field listening. I: stimulus synthesis. *J Acoust Soc Am* 85:858–867.
- Yin TC (2002) Neural mechanisms of encoding binaural localization cues in the auditory brainstem. In: *Integrative functions in the mammalian auditory pathway* (Oertel D, Fay RR, Popper AN, eds), pp 99–159. New York: Springer.
- Zonooz B, Arani E, Van Opstal AJ (2018) Learning to localise weakly-informative sound spectra with and without feedback. *Sci Rep* 8:17933.
- Zonooz B, Arani E, Kording KP, Aalbers PATR, Celikel T, Van Opstal AJ (2019) Spectral weighting underlies perceived sound elevation. *Sci Rep* 9:1642.
- Zwiers MP, Versnel H, Van Opstal AJ (2004) Involvement of monkey inferior colliculus in spatial hearing. *J Neurosci* 24:4145–4156.