Dynamic binaural sound localization based on variations of interaural time delays and system rotations

Claude Baumann,[1,a)] Chris Rogers,[2] and Francis Massen[1]
[1]*Lycée Classique Diekirch, Computarium, 32 Avenue de la Gare, Diekirch 9233, Luxembourg*
[2]*Tufts University, School of Engineering, 200 College Avenue, Medford, Massachusetts 02155, USA*

(Received 7 February 2015; revised 16 June 2015; accepted 22 June 2015; published online 6 August 2015)

This work develops the mathematical model for a steerable binaural system that determines the instantaneous direction of a sound source in space. The model combines system angular speed and interaural time delays (ITDs) in a differential equation, which allows monitoring the change of source position in the binaural reference frame and therefore resolves the confusion about azimuth and elevation. The work includes the analysis of error propagation and presents results from a real-time application that was performed on a digital signal processing device. Theory and experiments demonstrate that the azimuthal angle to the sound source is accurately yielded in the case of horizontal rotations, whereas the elevation angle is estimated with large uncertainty. This paper also proves the equivalence of the ITD derivative and the Doppler shift appearing between the binaurally captured audio signals. The equation of this Doppler shift is applicable for any kind of motion. It shows that weak binaural pitch differences may represent an additional cue in localization of sound. Finally, the paper develops practical applications from this relationship, such as the synthesizing of binaural images of pure and complex tones emitted by a moving source, and the generation of multiple frequency images for binaural beat experiments.
 [http://dx.doi.org/10.1121/1.4923448]

[ICB] Pages: 635–650

I. INTRODUCTION

One particularity of the auditory system in mammals is its physical proximity with the vestibular system, responsible for most of the animal balance and spatial orientation capacities. The interconnection of both different sensory systems suggests that motion somehow contributes to complex forms of analysis of the auditory scene and more particularly of sound localization. In this paper, the authors want to give evidence that the dynamics of head rotations play an important role in three-dimensional (3D) sound localization. The work starts from a reduced binaural system capable of sensing interaural time delays (ITDs) and angular velocity and uses a differential equation to describe the involved dynamics, which forms a special view on ITDs by considering their variability over time. It is shown that the knowledge of three quantities, the instantaneous ITD, its first derivative with respect to time and the rotation speed of the binaural system unambiguously and robustly yields the azimuth in the case of a horizontal rotation, thus solving the back/front confusion. Additionally and much more unintuitively, it is demonstrated that theoretically the magnitude of the elevation is also yielded. The work includes the analysis of error propagation and proves that the elevation can only be validly estimated if the sound source is not located in the horizontal or the median plane.

Furthermore, the paper presents results from a physical experiment of the model using a high-speed digital signal processing device. Finally, the authors demonstrate that one can determine the frequency modulation applied to the original acoustic signal because of the effect of head rotation. This might be considered as an additional cue in sound localization, expressing itself as a Doppler shift between the left and right signals. This binaural Doppler shift is proven to be mathematically equal to the ITD derivative.

II. ROTATIONS AND SOUND LOCALIZATION

A. Fundamentals

For an immobile binaural system, the knowledge of interaural time delays (ITDs) cannot solve the localization problem better than yielding a lateral surface of all possible locations of the sound source for a specific delay. This surface represents the sheet of a hyperboloid of confusion, which is generated by revolving one branch of the hyperbola that can be constructed for a certain value of the ITD upon two foci formed by the acoustic receivers.[1] Asymptotically, the hyperbolic sheet changes to a conical surface. The confusion in spatial localization of sound therefore only disappears if other cues are available.

In their mathematical and experimental study, Kneip and Baumann[2] have analyzed and proven the utility of deliberate rotations of a binaural system for the determination of the direction in space of a static sound source. The described active method is based on the measurements of ITDs before and after a single determined rotation of the interaural baseline about the z or y axis. The equations show that the angular component of the sound vector is unambiguously

[a)]Electronic mail: claude.baumann@education.lu

determined in the plane of rotation, whereas the orthogonal component appears in two symmetric instances. If the interaural axis is rotated about the z axis, as indicated on Fig. 1, the azimuth β is exactly yielded and the elevation ψ is found with ambiguity, which means that the magnitude of the elevation angle is known, but its sign is not. (Note that unlike the cited paper, this text uses the symbol ψ for the elevation. See Nomenclature.)

The present work starts from this viewpoint and deduces further equations from the properties of the cone of confusion. The following mathematical development proves that the combination of continuous measurements and differentiation of ITDs, and the measurements of the rotation speed in a given plane determines the angular component of the sound vector in that particular plane. Consequently, the magnitude of the second orthogonal component also is determined, whereas its sign is not. The disambiguation of this persistent uncertainty consists of clarifying whether the sound source is located in the upper/lower, or the frontal/backward hemispheres—depending on the rotation plane. This requires little additional information that can be easily obtained by complementary means, especially through a single subsequent orthogonal rotation.

The mathematical description restricts the analysis to a simple binaural sensor made of two omnidirectional microphones in the free field, augmented with a sensor for measuring angular velocity. Neither pinnae, nor head, nor torso are involved. Therefore, no spectral or level differences are considered. It is also assumed that the sound source, a single immobile spot in space, is located in the far field. According to Kneip and Baumann,[2] deviations between the asymptotic cone of confusion and the correct hyperbolic surface are less than 1° for sound distances greater than $4k$, and less than 0.1° for distances greater than $12k$, where $2k$ is the microphone spacing. Furthermore, the binaural system rotates about the z axis, while the speed of the microphone

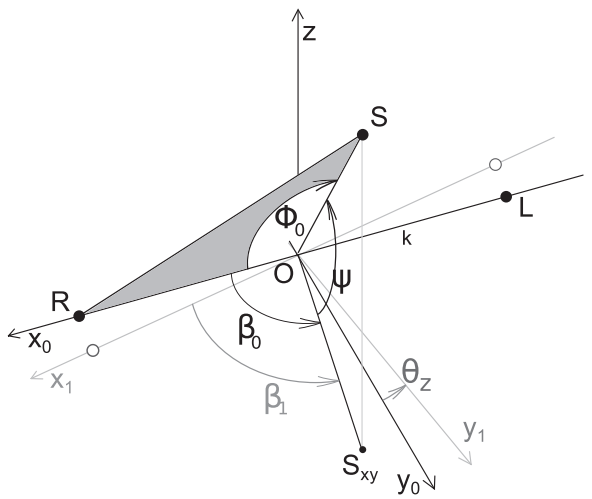

FIG. 1. Two microphones R and L, separated by distance $2k$, form a binaural system, which is rotated in the horizontal plane about its center O by angle θ_z. The sound source S and its associated horizontal projection S_{xy} yield the azimuth angles β_0 and β_1 in the coordinate systems (O, x_0, y_0) and (O, x_1, y_1). The elevation ψ is constant during rotations about the z axis. The angle ϕ_0 represents the direction to the sound source in the plane (R, L, S).

movements is very slow in comparison to the speed of sound. It must be underlined that the orthogonal system of coordinates (O, x, y, z) is always referenced to the binaural system using the interaural baseline as x axis. This must not be confused with the "world" reference frame. Because the coordinate system may be arbitrarily fixed in the median plane (O, y, z), it is obvious that the present study in fact covers any case of rotation of the binaural system, except the one about the x axis. It should be noted that the choice of this head referencing is consistent with the findings of Altmann *et al.*,[3] who concluded from their neurophysiological study that humans process spatial changes of sound source position in a head-related coordinate system.

This work is based on the insight that has been gained through the research about natural sound localization in combination with head rotations by van Soest,[4] who stated that through head movements, according to the phase theory, the sound direction is given by the intersectional line of successive cones of confusion. The approach of van Soest[4] has been verified by Reid and Milios[5] using an approximation method running on a digital signal processing device. Wallach[6] showed that head rotations provide essential information about the azimuth and the elevation of the sound source. His hypotheses have been testified by Perret and Noble[7] and Wightman and Kistler,[8] for instance. Thurlow *et al.*[9] observed three types of head rotations about each of the main axis that contribute to the localization of a sound source in space. Hill *et al.*[10] developed a simulation model in order to show that back/front ambiguity may be resolved through ITD variations arising from head rotation.

Lambert[11] developed a theory of sound-source localization in the horizontal plane using a function of the rate at which the interaural time delay changes with respect to the azimuthal angle. Rao and Xie[12] worked out a mathematical model, in order to prove that variations of ITDs caused by head rotations represent a valuable localization cue in the median plane (O, y, z). These authors used derivatives of the ITD with respect to changes in azimuth and altitude. The present work goes a step further by considering ITD derivatives with respect to time, as azimuth and altitude are themselves functions of time.

Recent research by Macpherson[13] confirms that in the absence of spectral cues, vestibular information that becomes available through head rotations represents the dominant cue for back/front localization. If the head is not allowed to move, interaural differences can only resolve the lateral localization. However, the temporal dynamics related to head rotations adds enough information to produce correct back/front disambiguation, if the stimulus is presented for more than ∼100 ms. Morikawa *et al.*[14] report that band-limited and band-unlimited white noise could be well localized for sources in the horizontal plane, if head movements were allowed. All subjects turned their heads in direction of the source, without necessarily centering it after the rotation. If the source was presented in the median plane (O, y, z), subjects yawed their heads left and right, independently of the elevation angle of the sound vector. Hirahara *et al.*[15] used the *Telehead*, a steerable dummy head, in order to verify the impact of head rotation on sound localization. The

Telehead was coupled to the listener's head through a motion sensor and followed the horizontal head rotation in real-time. Listener and *Telehead* were located in different experiment rooms, and the listener received sound stimuli from the *Telehead* via headphones. Results showed that voluntary head rotations play a crucial role in sound localization.

Blauert[16] underlines that motional theories of sound localization based on ITDs must necessarily be heterosensory, which directly follows from the present equations. Indeed, it is obvious that ITD detection alone can only discriminate the angle ϕ (cf. Fig. 1), which is the sound direction in the plane (R, L, S). Wallach[6] and Kneip and Baumann[2] showed that the angle ϕ represents a compound made of the azimuth β and the elevation ψ of the sound source. These three angles are related to each other through a variant on Wallach's formula,

$$\cos\phi = \cos\psi\cos\beta. \tag{1}$$

The ITD τ depends on the microphone spacing $2k$, the sound velocity c and the cosine of ϕ following a variant on von Hornbostel and Wertheimer's[17] equation using $\tau_{\max} = 2k/c$,

$$\tau = \tau_{\max}\cos\phi. \tag{2}$$

These important relations, which are directly connected to the definition of the cone of confusion, lead to the conclusion that ITD measurements performed by an immobile binaural system, incapable of monaural cues, can only discriminate the magnitude of the azimuth in the case of known elevation, or the magnitude of the elevation in the case of known azimuth. The identity $\phi = \pm\beta$, which follows for source locations in the horizontal plane has repeatedly been exploited, in order to conceive robots able to track sound sources in 2D. However, as the ambiguity of the sign is unresolved in that equation, the robots must somehow get additional information for correct back/front discrimination. For instance, Andersson *et al.*[18] solved the back/front ambiguity by using a spherical head. In addition to the measurements of interaural phase differences (IPD)—an alternative view upon temporal differences[19]—the device read interaural level differences (ILDs) during the motion toward the sound source. Portello *et al.*[20] used a binaural sensor made of two omnidirectional microphones in the free field. The processing device detected ITDs and applied a complex probabilistic algorithm, in order to have the robot track mobile sound sources in the horizontal plane. These two projects typify comparable attempts of using binaural systems in 2D sound localization.

Although it is impossible to yield the exact magnitude of the azimuth, if no other source of information is available than ITD measurements, the sign of the actual ITD value τ unambiguously determines in which lateral hemisphere the sound source is located. However, if the binaural system is allowed to move, a simple and efficient 2D sound localization strategy can be developed for a robot that additionally solves the back/front ambiguity (cf. Table I). Basically, this strategy exploits the lateral discrimination given by the sign of τ and rotates the robot accordingly, always ending in a stable steady state where the robot will face the sound source.

TABLE I. Simple 2D sound localization algorithm.

If $\tau > 0$ then rotate to the right
If $\tau < 0$ then rotate to the left
If $\tau = 0$ then stop rotating

There is a metastable special case that appears with this algorithm. A sound source that is initially located in the rear section of the median plane represents an unstable steady state. In such a condition, as ITDs actually are zero, the robot will stop rotating according to the algorithm. Yet, minimal perturbations will cause measurable non-zero ITDs and launch the rotational movement toward the stable steady state. Because the robot only reacts on lateral cues, it does not matter if the sound source is located in a different plane than the robot. The discussed algorithm is capable of physically localizing the horizontal sound direction by considering changes of the ITDs that are due to the robot's motion. One can evaluate variations of ITDs with respect to time by differentiating Eq. (2). The resulting equation, Eq. (3), is true at any time and for any kind of motion,

$$\dot{\tau}(t) = -\tau_{\max}\dot{\phi}(t)\sin\phi(t). \tag{3}$$

In the case of a horizontal rotation, the elevation remains invariant, and the derivative can be expressed with Eq. (4):

$$\dot{\tau}(t) = -\tau_{\max}\cos\psi\omega_{\beta}(t)\sin\beta(t). \tag{4}$$

This derivative is proportional to the angular speed $\omega_{\beta}(t) = \dot{\beta}(t)$ and the sine of the azimuth $\beta(t)$. The approach of using determined rotations and differentiating ITDs with respect to time for the sound localization problem introduces unexpected features. For instance, the back/front ambiguity may be solved without difficulty through the sign of the derivative for a known direction of rotation (cf. Table II). Note that this is true only if $\psi \neq \pm\pi/2$, where $\cos\psi > 0$. In the other case, the sound source is located on the axis of rotation, where no changes of τ can occur.

It must be underlined that the azimuth β is always expressed in the referential frame of the binaural system, as indicated in Fig. 1. The consequence of this referencing is expressed in Eq. (5) saying that a positive (anticlockwise)

TABLE II. Dynamically resolve the frontal ambiguity.

Perform a positive horizontal rotation of the binaural system in the "world" reference
If $\dot{\tau} > 0$ then source is located in the frontal hemisphere
If $\dot{\tau} < 0$ then source is located in the backward hemisphere
If $\dot{\tau} = 0$ then source is located on the frontal plane (O, x, z)

rotation of the system corresponds to the negative summation of the rotation speed:

$$\theta_z = -\int_{t_0}^{t_1} \omega_\beta(t)\mathrm{d}t, \tag{5}$$

during which time it moved from β_0 to β_1, so

$$\beta_1 = \beta_0 - \theta_z. \tag{6}$$

This becomes more evident through the following example. Assuming that the sound source is located at $\beta_0 = 45°$ and the rotation speed of the world frame is positive $\dot{\beta}_{\text{world}} = 5°/\text{s}$ ($\Longleftrightarrow \dot{\beta}_{\text{robot}} = -5°/\text{s}$). After 2 seconds, the system frame has rotated anticlockwise by 10°. Observed from the perspective of the binaural system, the sound source has moved clockwise. Therefore the new azimuth is $\beta_1 = 35°$. Also, the corresponding ITD values yield $\tau_1 > \tau_0$, because the source has come closer to the right microphone. Hence, the ITD derivative must be positive in that case.

For the case that $\tau(t) \neq 0$ and the elevation $\psi \neq \pm(\pi/2)$, it follows that the relative variation of ITDs may be expressed as [combining Eqs. (2), (1) and (4)]

$$\frac{\dot{\tau}(t)}{\tau(t)} = -\omega_\beta(t)\tan\beta(t). \tag{7}$$

In other words, the relative variation of ITDs is proportional to the angular speed of the interaural axis in the horizontal plane and the tangent of the azimuth. Equation (7) proves that it is possible with the knowledge of the instant ITD, its derivative and the rotation speed, to unambiguously determine the azimuth at any moment. Because the sign of τ resolves the lateral ambiguity, Eq. (7) delivers a unique value of β for any valid combination of $\dot{\tau}$, τ, and ω_β. Notably, the equation is independent of the sound velocity and the microphone spacing.

It might be interesting to analyze the case, where the system moved from β_0 to β_1 by integrating Eq. (7) over time (disregarding the discussion about signs and domains),

$$\int_{t_0}^{t_1} \frac{\dot{\tau}(t)}{\tau(t)}\mathrm{d}t = \int_{t_0}^{t_1} -\dot{\beta}(t)\tan\beta(t)\mathrm{d}t \tag{8}$$

Using Eq. (6) and defining $\tau_{0,1} = \tau(t_{0,1})$, Eq. (8) yields (cf. Appendix A)

$$\tan\beta_1 = \cot\theta_z - \frac{\tau_0}{\sin\theta_z\tau_1}. \tag{9}$$

Therefore, if one measures two ITDs (τ_0 and τ_1) at some arbitrary time difference and the angle θ_z through which the microphones have moved in that time, one can estimate the azimuthal location of the source β_1 relative to the microphone axis at the end time of the rotation. This equation [Eq. (9)] has been worked out geometrically in the cited work by Kneip and Baumann.[2]

B. Uncertainty analysis

The present paper does not discuss techniques for estimating ITDs and rotation speed. ITD derivatives may be obtained by applying appropriate numerical differentiation methods in the case of digital signal processing. In such a case, the relative uncertainty of the ITD derivative can be written

$$\frac{\Delta\dot{\tau}}{|\dot{\tau}|} = \sqrt{\left(\frac{\Delta\tau}{\tau}\right)^2 + \left(\frac{\Delta t}{t}\right)^2}. \tag{10}$$

As it may be supposed that the error in timing is negligible, the uncertainty of the derivative may be expressed with

$$\Delta\dot{\tau} = |\dot{\tau}|\frac{\Delta\tau}{|\tau|}. \tag{11}$$

1. Uncertainty in azimuthal measurements

Rewriting Eq. (7),

$$\beta = \arctan\left(-\frac{\dot{\tau}}{\tau\omega_\beta}\right),\ \tau \neq 0,\ \omega_\beta \neq 0, \tag{12}$$

and using Eqs. (11) and (12), the error propagation in azimuth can be expressed applying the first order Taylor expansion over several stages,[21]

$$\begin{aligned}\Delta\beta &= \left|\frac{\partial\beta}{\partial\dot{\tau}}\right|\Delta\dot{\tau} + \left|\frac{\partial\beta}{\partial\tau}\right|\Delta\tau + \left|\frac{\partial\beta}{\partial\omega_\beta}\right|\Delta\omega_\beta \\ &= \left|\frac{2\sin\beta}{\cos\psi}\right|\left|\frac{\Delta\tau}{\tau_{\max}}\right| + \left|\frac{\sin 2\beta}{2}\right|\left|\frac{\Delta\omega_\beta}{\omega_\beta}\right|.\end{aligned} \tag{13}$$

Figure 2 shows the influence of the elevation on the uncertainty in azimuthal angle for typical values of the involved parameters. The authors used a relative uncertainty in ω_β of 10% and in the ITD of 2%. The uncertainty in azimuthal angle is smallest for sound sources in the horizontal plane. Uncertainty in ITDs have a minor influence on azimuth estimations. The uncertainty dramatically grows as the rotation speed slows to zero. This is not visible in Fig. 2 but directly follows from Eq. (13).

2. Uncertainty in elevation angle

At any moment, the knowledge of the azimuth β allows the determination of the absolute value of the elevation ψ,

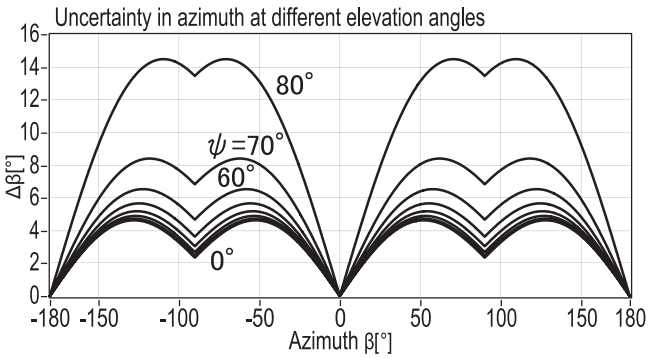

FIG. 2. The uncertainties in azimuth $\Delta\beta$, expressed in degrees, depend on the azimuth and the elevation. In this graph the relative uncertainty in rotation speed $\Delta\omega_\beta/\omega_\beta = 10\%$ and in ITD $\Delta\tau/\tau_{\max} = 2\%$.

independently of any motion of the binaural system. However, the equations cannot discriminate whether the sound source is located in the upper or lower hemisphere. Disregarding the missing sign, uncertainty bounds in elevation can be estimated by combining Eqs. (1) and (2), in order to form Eq. (14),

$$\psi = \arccos\left(\frac{\tau}{\tau_{\max}\cos\beta}\right). \tag{14}$$

The uncertainty in elevation is calculable through Eqs. (13) and (14),

$$\begin{aligned}\Delta\psi &= \left|\frac{\partial\psi}{\partial\tau}\right|\Delta\tau + \left|\frac{\partial\psi}{\partial\beta}\right|\Delta\beta \\ &= \left|\frac{1}{\sin\psi\cos\beta}\right|\frac{\Delta\tau}{\tau_{\max}} + |\cot\psi\tan\beta|\Delta\beta.\end{aligned} \tag{15}$$

Figure 3 shows the uncertainty in elevation in function of the azimuth. Errors strongly depend on the elevation, and rapidly grow, as β approaches $\pm\pi/2$. Uncertainty in elevation is smallest for sound sources that are located on the frontal plane, and greatest for sound sources located near the horizontal or the median planes.

C. Experimental applications

In order to study the practical value of the described dynamic localization model, the equations have been implemented into a digital signal processing device with real-time capacities. The binaural sensor was composed of two omnidirectional Electret condenser microphones fixed on a rod and two high gain audio amplifiers (cf. Fig. 4). Spacing $2k$ between the microphones was 19.5 cm. The binaural sensor was mounted on a motorized structure, which was horizontally rotatable at variable angular speed between 0 and π rad/s. A rotary encoder measured angular velocity and position for control purposes at 0.5° resolution.

Experiments were performed in semi-anechoic room conditions (empty cinema theater with unknown attenuation coefficient) at different elevation angles. Sound waves were emitted by a small music playing radio at a distance of 2 m ($>12k$) from the microphones. The digital signal acquisition and processing device consisted of the National Instruments

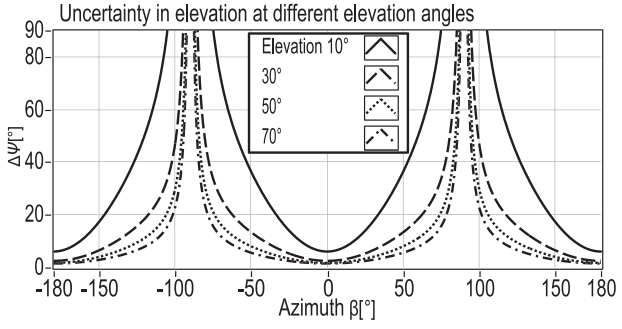

FIG. 3. This graph evaluates the uncertainty in elevation $\Delta\psi$ in function of the azimuth at different elevation angles. (Identical conditions as explained in Fig. 2.)

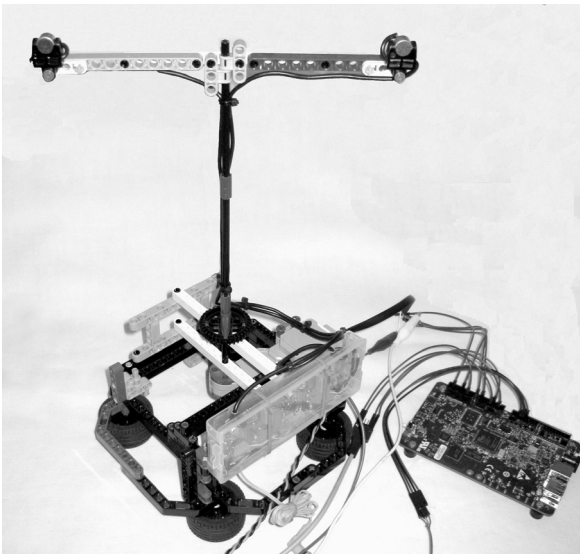

FIG. 4. The picture shows the steerable binaural system made of LEGO™ parts, microphones, amplifiers and uncased FPGA-based digital signal processing (DSP) device.

(Austin, TX) myRIO™ embedded hardware device, a dual-core real-time system based on a customizable field programmable gate array (FPGA).

ITDs were extracted from digitized audio signals by finding the peak in the running cross-correlation function, as proposed by Sayers and Cherry.[22] In contrast to the method applied by these authors, a sliding rectangular time-window of 10 ms was used instead of the original exponential window. Because timing was critical, especially during fast rotations of the binaural system, the normalized cross-correlation function was calculated in the time domain applying a fast recursive algorithm[23] that allowed to produce ITD at the audio sampling frequency $f_s = 20$ kHz. Also, the resolution in ITD measurement was $1/f_s$. In order to eliminate bad cues, ITD values were prefiltered with an adaptive digital filter using the degree of coherence as a weight (cf. Blauert,[16] p. 201). The windowing and prefiltering was about 40 ms. In the case of a rotation speed of 1 rad/s, this signified an additional uncertainty less than 2.5° in appreciation of the angle ϕ. Such delay effects in localization have been observed with human listeners. The existence of binaural sluggishness suggests that ITD appreciation requires some processing time in natural hearing as well. Grantham and Wightman[24] evaluate this delay between 44 and 243 ms.

ITDs were differentiated using a first-order polynomial low-pass filter that was especially designed for the purpose. Numerical differentiation requires particular precaution, as the involved operations inevitably introduce truncation and round off errors that may become excessively high.[25] Moreover, because the underlying ITD function $\tau(t)$ is unmodeled, error-prone and susceptible to noise, simple methods for computing the numerical derivative were not helpful. Therefore, a least-square smoothing algorithm[26] was used, in order to guarantee sufficient noise suppression, low distortion and zero-phase characteristics in the relevant angular frequency band. This feature was essential, in order to prevent that the filtering process would generate malign

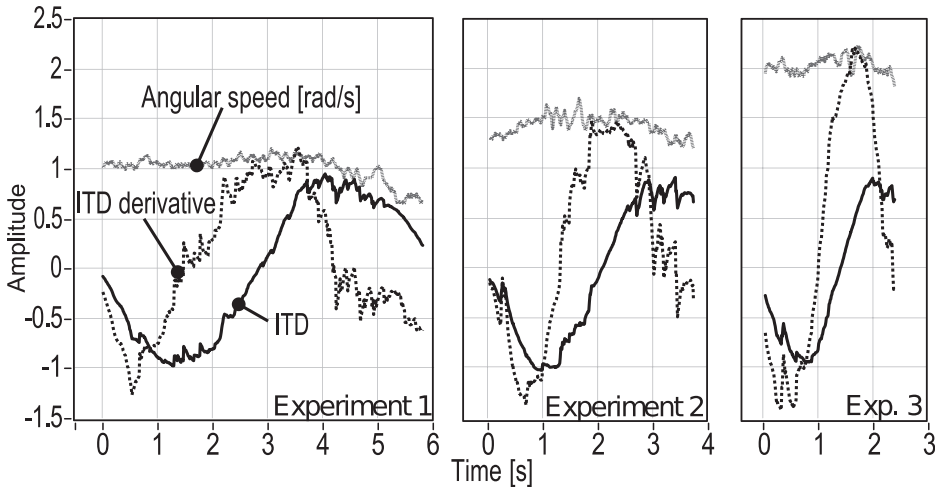

FIG. 5. This graph depicts three exemplary experiments. ITDs are expressed as relative values τ/τ_{max}, and the derivatives as $\dot{\tau}/\tau_{max}$. Angular speed is referenced to the binaural coordinate system. Therefore, the sign is positive, although the rotation of the binaural system has been performed clockwise.

phase shifts of the ITD derivative $\dot{\tau}(t)$ with respect to $\tau(t)$, which would cause misinterpretations of sound direction.

Applying Gauss' method of the least-squares, under the assumption of constant time steps Δt, a synchronized estimation at current time $t_i = i\Delta t$ of the derivative is given by the following discrete polynomial function $p[i]$ for M consecutive past values of the difference quotient $q[j] = (\tau[j] - \tau[j-1])/\Delta t$ (cf. Appendix E),

$$p[i] = \frac{\sum_{j=0}^{M-1} (6j - 2(M-2))q[j]}{M(M+1)}. \tag{16}$$

In order to reduce the number of computations for large values of M, the filter was modified as follows.

The function, Eq. (16), can be rewritten,

$$p[i] = \frac{6S_{iq}[i] - 2(M-2)S_q[i]}{M(M+1)}, \tag{17}$$

where at any time index i the summations S_q and S_{iq} are computed upon the latest M values of q,

$$S_q[i] = \sum_{j=0}^{M-1} q[j] \text{ and } S_{iq}[i] = \sum_{j=0}^{M-1} j(q[j]). \tag{18}$$

These sums can be calculated easily and independently of M (in terms of computing time) with Eq. (19), which at any time index i, exclusively considers the effects of the departure of the oldest and the arrival of the most recent values of the difference quotient stored in an M-sized first-in, first-out (FIFO) data buffer with circular index,

$$\begin{aligned} S_q[i] &= S_q[(i-1) \bmod M] - q_{\text{old}}[i] + q_{\text{new}}[i], \\ S_{iq}[i] &= S_{iq}[(i-1) \bmod M] - S_q[i] + M(q_{\text{new}}[i]). \end{aligned} \tag{19}$$

(Note that the correctness of this algorithm can be demonstrated without difficulty by induction.)

Figure 5 shows typical filtered measurements of ITD, ITD derivative, and rotation speed. The measurements were made at elevation $\psi = 0$. The robotic system was programmed to perform a full revolution of the binaural baseline at different angular speeds. Figure 6 displays the results of the estimations that have been made by applying the model equations of this paper. Visibly, the azimuth is well estimated in comparison with the measured position [standard deviation in azimuth is $\sigma_{(\beta,\psi=0)} = 0.2$ rad]. By contrast, although expectedly, because of the propagation of errors, the elevation of $\psi = 0$ is estimated in the three experiments with mean $\mu_1 = 0.6$ rad (standard deviation, $\sigma_1 = 0.3$), $\mu_2 = 0.4$ rad ($\sigma_2 = 0.4$), and $\mu_3 = 0.3$ rad ($\sigma_3 = 0.4$). A further experiment yielded for

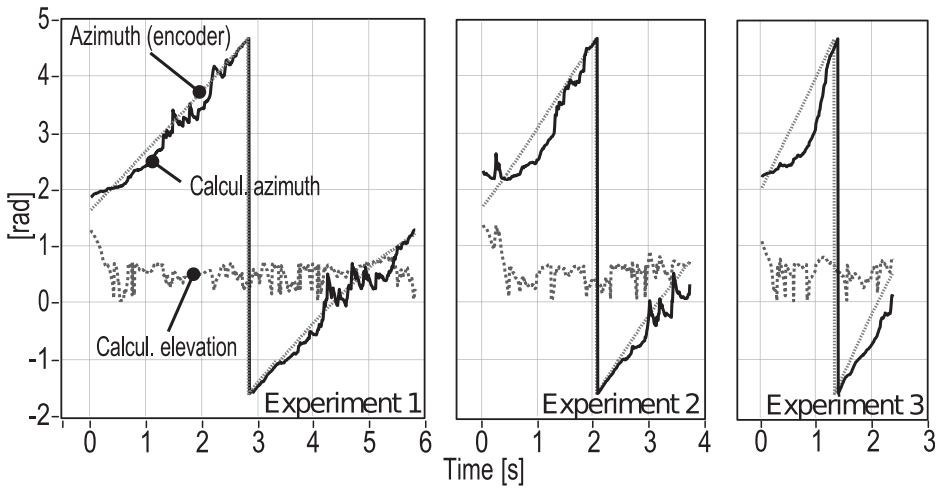

FIG. 6. This graph compares the calculated azimuth values with the reference values recorded by the rotary encoder. As expected for sound sources in the horizontal plane, elevation estimations present large errors during most of the time.

$\psi = 0.5$ rad, $\sigma_{(\beta,\psi=0.5)} = 0.33$ rad, and $\mu = 0.6$ rad ($\sigma = 0.36$) in elevation, for instance.

In conclusion, the experiments were necessary to find out that the critical part of the application of the present model is related to the synchronicity of the ITD function $\tau(t)$ and its derivative $\dot{\tau}(t)$, because common methods of numerical differentiation of noisy signals introduce an undesired delay between both functions. The results of the tests confirmed that the method yields excellent estimations of the azimuth in the case of horizontal rotations with growing errors as the elevation is increased. The method provides rather bad estimations of the elevation, which improve with rising elevation at the cost of the azimuthal measurement.

III. MOTION AND BINAURAL DOPPLER EFFECT

In the context of binaural hearing, the appearance of the Doppler effect in combination with head movements or motion of the sound source in the horizontal plane has been studied by Jenison,[27] who listed first order formulas showing that source position, velocity and distance are observable from combined measurements of interaural time delay, Doppler effect and sound intensity. Müller and Schnitzler[28] stepped further into the subject by trying to find a relationship between the interaural time delays and the Doppler effect. The authors studied the Doppler effect on the head considered as a single spot rather than on the binaural system with two receivers. Neelon and Jenison[29] focused on the comparison of translatory and rotational source movements. From observation they concluded that a source that is rotating around a binaural system does not produce any noticeable Doppler shift. The present paper will prove evidence that the Doppler shift exists in spite of these results. Only it is very weak, in the case of slow rotations, as applied by Neelon and Jenison.[29] Iwaya *et al.*[30] developed a rather complex algorithm based on head-related impulse responses (HRIRs) that adds the Doppler shift to the computations of time delays. Kumon and Uozumi[31] proposed an auditory system using an extended Kalman filter (EKF), which fuses the motion of a mobile robot platform with the sensory information of a binaural ITD detector carried by the robot. The applied model included the Doppler shift in the calculations.

In Sec. III A, three special cases of motion will be studied: (1) rotations of the binaural system (with a stationary sound source); (2) rotations of the sound source about the center (with a stationary binaural system); and (3) translations of the sound source in parallel to the binaural baseline (again with the binaural system staying stationary). Finally, (4) the general case of an arbitrary trajectory of the sound source will be investigated. The present developments will prove that the Doppler shift between the binaurally perceived waves and the first derivative of the interaural time delay are equivalent. It must be recognized that motion is described in the plane (R, L, S) and that the sound distance is greater than $12k$, so that the true sound direction coincides with the slope of the asymptote of the related hyperbola of confusion ($\phi \simeq \varphi$) (cf. Fig. 8).

A. Theory

1. Rotating binaural system, stationary source

Assume that a stationary sound source is emitting a continuous 100 Hz sine-wave and is located in the far field at $\beta = 0$. The corresponding ITD $\tau_{t=0} = \tau_{\max}$ is made visible in Fig. 7 through the apparent delay between the left and right signals. After a rapid rotation at $\sim$30 rad/s, the source is located at $\beta = \pi$ rad with $\tau_{t=0.1} = -\tau_{\max}$. The picture shows that now the signal recorded on the right microphone lags behind the left one. The question arises, how both signals, originated from the same source change in shape and timing, in order to perform the time shift during head motion. Because the binaural system is totally independent of the sound source, it has no influence on the sound generating process. It continuously receives two complementary sound signals that are both similar and distinctive. Their frequencies at arrival must change, although the frequency is constant at the source. This Doppler shift with a stationary source and rotating receivers is due to the variations of the pressure wave arrival times to each microphone.

Figure 8 outlines the plane containing the binaural system and the sound source. The three relevant distances d_R, d_L, d are unknown. If the binaural system rotates about any axis passing through the center of the binaural system O, the sound distance d does not change. However, the angle ϕ, and subsequently all other involved variables become functions of time: $\phi(t)$, $d_R(t)$, $d_L(t)$, etc. Note that for better readability the reference to time will be subsequently dropped, although it is implicit for all of the variables.

Following the development of Kneip and Baumann[2] (p. 3110), both microphones R and L that form the binaural system are the focuses of a hyperbola of confusion in the plane

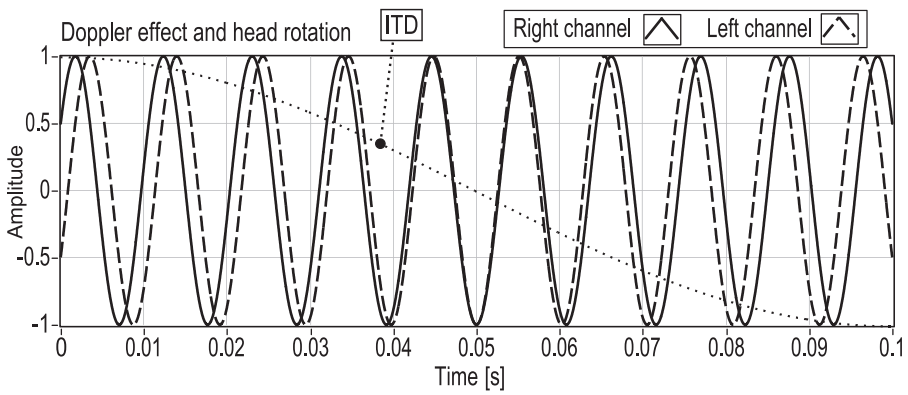

FIG. 7. The right and left signals are both exposed to frequency modulation because of the Doppler effect. The thin dotted line represents the ITD function expressed as $\tau/\tau_{\max}$.

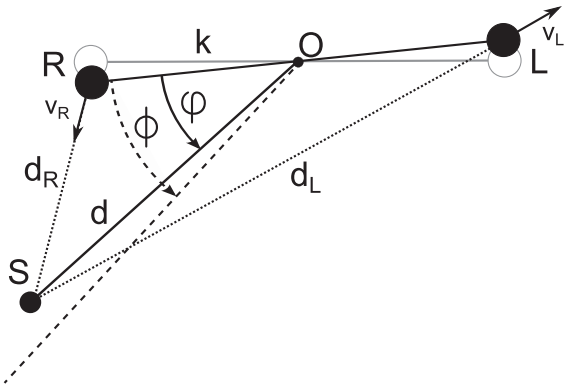

FIG. 8. A rotation of the binaural system in the plane (R, L, S) with the involved distances and radial speeds. The picture emphasizes the true sound direction φ and the slope of the cone of confusion ϕ, which coincide for sound sources in the far field.

(R, L, S), and the angle ϕ representing the slope of the characteristic asymptote of that hyperbola is defined by

$$\cos\phi = \frac{d_L - d_R}{2k} = \frac{c\tau}{2k} \tag{20}$$

with

$$\tau = \frac{d_L - d_R}{c}. \tag{21}$$

The variables may be differentiated with respect to time,

$$\dot{\tau} = \frac{1}{c}\left(\frac{\mathrm{d}}{\mathrm{d}t}(d_L) - \frac{\mathrm{d}}{\mathrm{d}t}(d_R)\right). \tag{22}$$

If d_L or d_R are decreased, the respective microphone verges on the sound source. Therefore, let

$$\begin{aligned} v_L &= -\frac{\mathrm{d}}{\mathrm{d}t}(d_L), \\ v_R &= -\frac{\mathrm{d}}{\mathrm{d}t}(d_R). \end{aligned} \tag{23}$$

(It is assumed here that, depending on the direction of motion on the radial axes, $v_L = \pm\|\vec{v}_L\|$ and $v_R = \pm\|\vec{v}_R\|$.)

Finally, combining these equations results in

$$\dot{\tau} = \frac{v_R - v_L}{c}. \tag{24}$$

Assuming that a pure tone is emitted by the source with frequency f_A, the wavelengths are the same at the source λ_A and at the microphones (λ_R and λ_L), because the sound source is stationary in the fluid medium. However, as the microphones are moving, the speed of sound is virtually increased (or decreased) by the speeds v_R and v_L, the frequencies f_R and f_L must change accordingly,

$$\begin{aligned} &\lambda_R = \lambda_A, \\ &\frac{(c + v_R)}{f_R} = \frac{c}{f_A} \\ &\Rightarrow f_R = \left(1 + \frac{v_R}{c}\right) f_A \\ &\text{and similarly } f_L = \left(1 + \frac{v_L}{c}\right) f_A. \end{aligned} \tag{25}$$

Subtracting these equations leads to an expression of the binaural Doppler shift,

$$\frac{f_R - f_L}{f_A} = \frac{1}{c}(v_R - v_L) = \dot{\tau}. \tag{26}$$

The role of the unknown source frequency f_A in this equation is to set up the magnitude of the frequency shift.

The law of cosines yields

$$\begin{aligned} d_R &= \sqrt{A - B\cos\varphi} \simeq \sqrt{A - B\cos\phi}, \\ d_L &= \sqrt{A + B\cos\varphi} \simeq \sqrt{A + B\cos\phi}, \\ A &= k^2 + d^2, \\ B &= 2kd, \end{aligned} \tag{27}$$

$$\begin{aligned} &\frac{\mathrm{d}}{\mathrm{d}t}(d_R) = \zeta_R \dot{\phi}(t)\sin\phi(t) \text{ with} \\ &\zeta_R = \frac{kd}{d\sqrt{1 + \frac{k^2}{d^2} - \frac{2k}{d}\cos\phi(t)}}. \end{aligned} \tag{28}$$

Important note: Eq. (28) is only true, because k and d are constants,

$$\Rightarrow \lim_{d\to\infty} \zeta_R = k. \tag{29}$$

Therefore, for sufficiently large d compared to k,

$$v_R = -k\dot{\phi}(t)\sin\phi(t), \tag{30}$$

and similarly

$$v_L = +k\dot{\phi}(t)\sin\phi(t), \tag{31}$$

whence

$$v_L = -v_R. \tag{32}$$

Applying Eq. (25),

$$f_R + f_L = f_A\left(2 + \frac{v_R + v_L}{c}\right) = 2f_A \Rightarrow f_A = \frac{f_R + f_L}{2}. \tag{33}$$

Finally, by combining Eqs. (26) and (33), the derivative of the ITD function may be written as the binaural Doppler equation,

$$\dot{\tau} = 2\frac{f_R - f_L}{f_R + f_L}. \tag{34}$$

In conclusion, rotations of the binaural system in any plane produce relative changes of frequency that are equal in all frequency band for a source away from the microphones.

2. Rotating sound source, stationary microphones

The development of the binaural Doppler equation from Sec. III A 1 needs to be reconsidered here, because the

conditions are not the same in the case of a moving sound source and immobile binaural system (cf. Fig. 9). The wavelengths at the source λ_A and the receivers λ_R and λ_L are no longer identical, because the source is moving in the fluid medium. For instance, if the source is moving away from the right receiver at speed $v_{SR} = -v_R$, the wavelength of the signal generated at the source is lengthened by the traveled distance within the duration of one cycle (note that $v_{SL} = -v_L$):

$$\lambda_R = \lambda_A - \frac{v_R}{f_A}, \quad \lambda_L = \lambda_A - \frac{v_L}{f_A} \Rightarrow f_R = f_A\left(1 - \frac{v_R}{c}\right)^{-1},$$
$$f_L = f_A\left(1 - \frac{v_L}{c}\right)^{-1}. \tag{35}$$

If the involved speeds are much smaller than the speed of sound, the binomial series approximation can be applied,

$$(1+q)^\alpha \approx 1 + \alpha q, \text{ if } q \ll 1$$
$$\Rightarrow f_R - f_L = f_A\left(1 + \frac{v_R}{c} - \left(1 + \frac{v_L}{c}\right)\right). \tag{36}$$

Because the development of Eqs. (27) to (33) can be applied to a rotating sound source and stationary microphones, the Doppler shift obeys the same equations as discussed in Sec. III A 1,

$$\frac{f_R - f_L}{f_A} = \frac{1}{c}(v_R - v_L) = \dot{\tau}, \tag{37}$$

which leads to

$$\Rightarrow \dot{\tau} = 2\frac{f_R - f_L}{f_R + f_L}. \tag{38}$$

In conclusion, both cases of rotation, either with rotating auditory system, or with rotation of the source, produce the same Doppler shift.

3. Shifting sound source, stationary microphones

The present section will analyze the case, where a source is moving linearly at constant speed $|v_x| \ll c$ in parallel to the binaural system, as shown in Fig. 10.

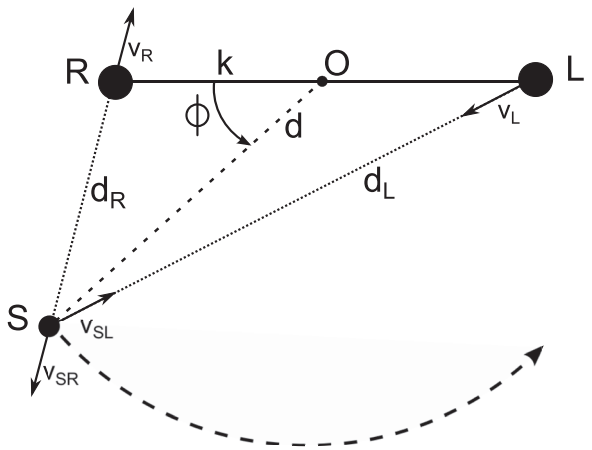

FIG. 9. The sound source is rotating around the binaural system in the plane (R, L, S).

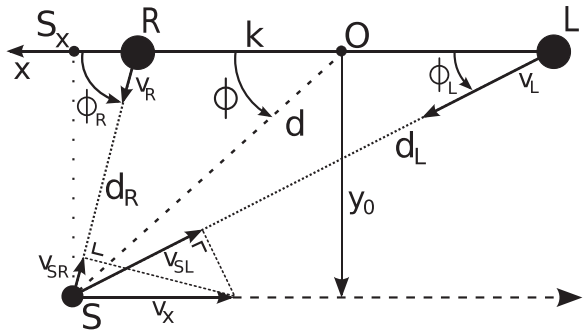

FIG. 10. The sound source is moving in parallel to the binaural system in the plane (R, L, S).

The motion of the sound source is described through the equations

$$\begin{cases} x = x_0 + v_x t = d\cos\phi \\ y = y_0. \end{cases} \tag{39}$$

Considering the triangles (R, S_x, S) and (L, S_x, S), the radial speeds can be expressed with

$$v_R = -v_{SR} = -v_x \cos\phi_R = -v_x \frac{x-k}{d_R},$$
$$v_L = -v_{SL} = -v_x \cos\phi_L = -v_x \frac{x+k}{d_L}. \tag{40}$$

Because for sufficiently large sound distances, Eqs. (20) and (27) can be applied,

$$d_L - d_R = 2k\cos\phi,$$
$$d_L + d_R = \frac{(d_L + d_R)(d_L - d_R)}{d_L - d_R} = \frac{d_L^2 - d_R^2}{2k\cos\phi}$$
$$\simeq \frac{k^2 + d^2 + 2kd\cos\phi - k^2 - d^2 + 2kd\cos\phi}{2k\cos\phi}$$
$$= 2d. \tag{41}$$

Using Appollonius' theorem of the median,

$$d_L^2 + d_R^2 = 2(k^2 + d^2)$$
$$\Rightarrow d_L d_R = \frac{(d_L + d_R)^2 - 2(k^2 + d^2)}{2} \approx d^2 - k^2. \tag{42}$$

Therefore,

$$v_R + v_L = -v_x \frac{d_L(x-k) + d_R(x+k)}{d_L d_R}$$
$$= -v_x \frac{(d_L + d_R)x - (d_L - d_R)k}{d_L d_R}$$
$$= -v_x \frac{2d^2\cos\phi - 2k^2\cos\phi}{d^2 - k^2}$$
$$= -2v_x\cos\phi \tag{43}$$

and

$$f_R + f_L = f_A\left(2 + \frac{v_R + v_L}{c}\right) = 2f_A\left(1 - \frac{v_x\cos\phi}{c}\right). \tag{44}$$

Hence

$$f_A = \left(1 - \frac{v_x \cos\phi}{c}\right)^{-1} \frac{f_R + f_L}{2}. \tag{45}$$

In this equation, v_x is supposed to be small in comparison to c, so that the first factor can be considered as 1. Because Eq. (37) is applicable here, the Doppler equation becomes

$$\dot{\tau} = 2\frac{f_R - f_L}{f_R + f_L}. \tag{46}$$

In conclusion, slow parallel shifts of the binaural system obey the same binaural Doppler rule as rotations.

4. Arbitrary trajectory of the sound source, stationary microphones

The motion of the sound source on an arbitrary trajectory in the plane (R, L, S) with speed vector $\vec{v} = \vec{v}_x + \vec{v}_y$ and $\|\vec{v}\| \ll c$ can be described with (cf. Fig. 11) the following:

$$\begin{aligned} x &= d\cos\phi, \quad \dot{x} = v_x, \\ y &= d\sin\phi, \quad \dot{y} = v_y, \end{aligned} \tag{47}$$

where, depending on the direction of motion on the x or y axis,

$$v_x = \pm\|\vec{v}_x\| \quad \text{and} \quad v_y = \pm\|\vec{v}_y\|. \tag{48}$$

The distance to the sound source is not necessarily constant, so (cf. Appendix B),

$$\begin{aligned} d &= \sqrt{(x^2 + y^2)}, \\ \dot{d} &= v_x\cos\phi + v_y\sin\phi, \\ \phi &= \arctan\left(\frac{y}{x}\right), \\ \dot{\phi} &= \frac{v_y\cos\phi - v_x\sin\phi}{d}. \end{aligned} \tag{49}$$

The distances to the microphones are defined by

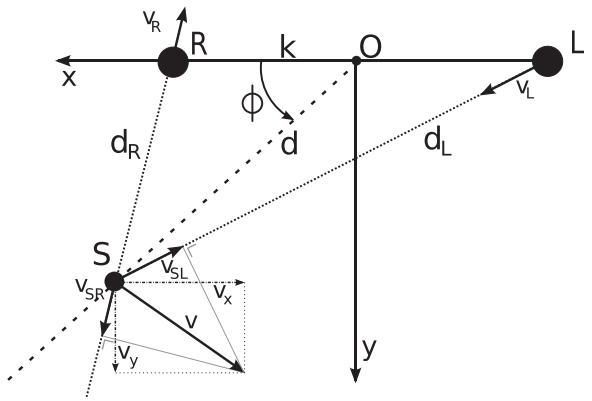

FIG. 11. The sound source is moving into an arbitrary direction in the plane (R, L, S).

$$\begin{aligned} d_R &\simeq \sqrt{A - B\cos\phi}, \\ d_L &\simeq \sqrt{A + B\cos\phi} \end{aligned} \tag{50}$$

with

$$\begin{aligned} A &= k^2 + d^2, \\ B &= 2kd \end{aligned} \tag{51}$$

and

$$\begin{aligned} \dot{A} &= 2d\dot{d}, \\ \dot{B} &= 2k\dot{d}. \end{aligned} \tag{52}$$

The radial speed v_R may be expressed with (cf. Fig. 11 and Appendix C)

$$\begin{aligned} v_R &= -v_{SR} = -\dot{d}_R \\ &= -\frac{d\dot{d} - kv_x}{d_R}. \end{aligned} \tag{53}$$

And similarly,

$$v_L = -\frac{d\dot{d} + kv_x}{d_L}. \tag{54}$$

Therefore, applying Eq. (41) and Appolonius' theorem [cf. Eq. (42) and Appendix D]

$$v_R + v_L \simeq -2(v_x\cos\phi + v_y\sin\phi). \tag{55}$$

Finally, and similarly to Eq. (45),

$$f_A = \left(1 - \frac{v_x\cos\phi + v_y\sin\phi}{c}\right)^{-1} \frac{f_R + f_L}{2} \simeq \frac{f_R + f_L}{2}, \tag{56}$$

and from Eq. (37),

$$\dot{\tau} = 2\frac{f_R - f_L}{f_R + f_L}. \tag{57}$$

In conclusion, the binaural Doppler equation is applicable for any kind of motion of the sound source given the involved speeds are much smaller than the sound velocity.

B. Applications

1. Generating a binaural image

The Doppler equations developed so far describe the relationship between the binaurally received frequencies and the ITD derivative. These equations may serve for the simulation of motion in the context of binaural hearing.

Assume that a sound source in the far field is emitting a pure tone of constant angular frequency $\omega_A = 2\pi f_A$ and zero phase. The signal amplitude is 1,

$$s_A(t) = \sin\omega_A t. \tag{58}$$

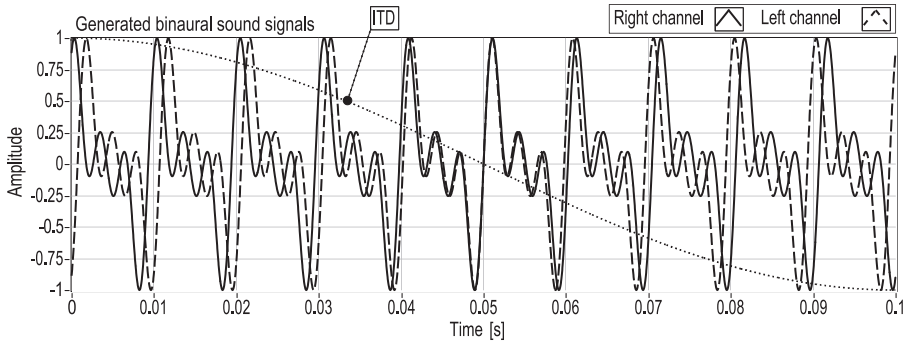

FIG. 12. The right and left signals have been generated by applying the equations of Sec. III B 1 to each component of a complex signal. Similarly to Fig. 7, the thin dotted line represents the ITD function expressed in $\tau/\tau_{\max}$.

Furthermore suppose that there are no signal losses and that the binaural system is rotating in the plane (R, L, S) with angular frequency $\omega_\phi(t) = \dot{\phi}(t)$. (Note that the function $\phi(t)$ may be an arbitrary continuous function of time.)

The generated signals are the results of complementary frequency modulations of the source signal. Applying Eqs. (25), (30), and (31), the instantaneous frequencies can be expressed with

$$\begin{aligned} f_R(t) &= \left(1 - \frac{k}{c}\omega_\phi(t)\sin\phi(t)\right) f_A \\ &= \left(1 - \frac{\tau_{\max}}{2}\omega_\phi(t)\sin\phi(t)\right) f_A = \left(1 + \frac{\dot{\tau}}{2}\right) f_A, \\ f_L(t) &= \left(1 - \frac{\dot{\tau}}{2}\right) f_A. \end{aligned} \tag{59}$$

According to the rules of frequency modulation (cf. Hartmann[32]), the signal received at the right microphone changes to

$$\begin{aligned} s_R(t) &= \sin\left(\omega_A \int_0^t \left[1 + \frac{\dot{\tau}}{2}\right] \mathrm{d}t\right) \\ &= \sin\left(\omega_A \left[t + \frac{\tau}{2}\right]\right) = \sin\left(\omega_A \left[t + \frac{\tau_{\max}}{2}\cos\phi(t)\right]\right). \end{aligned} \tag{60}$$

Note that the integration constant has been dropped.

Similarly,

$$\begin{aligned} s_L(t) &= \sin\left(\omega_A \left[t - \frac{\tau}{2}\right]\right) \\ &= \sin\left(\omega_A \left[t - \frac{\tau_{\max}}{2}\cos\phi(t)\right]\right). \end{aligned} \tag{61}$$

In the case of a constant angular frequency ω_ϕ, the sound direction changes according to the equation $\phi(t) = \phi_0 + \omega_\phi t$. Figure 12 shows binaural signals that have been synthesized by applying Eqs. (60) and (61) to each component of a complex sound signal.

In conclusion, because any desired physical motion of the sound source can be analytically transformed into variations of the compound angle ϕ, the developed frequency and phase equations may be applied for the generation of binaural sound signals. If the constructed sound signals are presented to a subject via headphones, it is likely that the impression of source motion will be perceived. However, as neither back/front nor up/down ambiguity is resolved, only lateralization will be possible (cf. Plenge[33]).

2. Constructing a binaural beat

The important relationship described in Eq. (34) may be illustrated through an experiment known in psychoacoustics as the binaural beat (cf. Rayleigh,[34] Licklider *et al.*,[35] and Akeroyd[36]). A typical form of the binaural beat consists in presenting a 500 Hz sine wave to one ear and a 501 Hz sine wave to the other (first curve in Fig. 13). Because the frequency difference is very small, humans do not perceive two separate tones. Instead a single tone is heard, which seems to move across the head at a rate of 1 Hz from one side to the other, then flip back, and restart its virtual motion. Because the appreciation of direction reaches its extrema $\phi = 0, \pi$ at $\tau_{\max}$ and $-\tau_{\max}$, respectively, ITDs have been clipped in Fig. 13.

This perception of motion directly follows from the binaural Doppler shift equation Eq. (34). In fact, the time lag between both signals (=ITD), induced by the slightly shorter wavelength of one signal, increases at a constant rate, until both signals are in phase again.

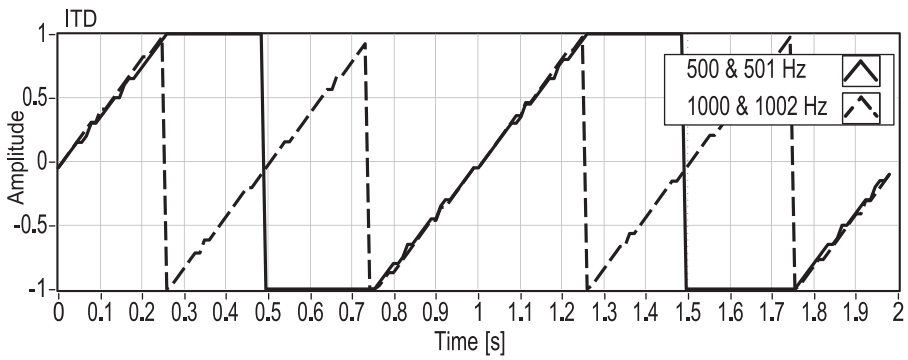

FIG. 13. Binaural beat experiment. Two pure tones were generated with frequencies 500 and 501 Hz (1000 and 1002 Hz). Each signal was recorded at the opposite input of the binaural system. ITDs were calculated through the running cross-correlation method. The graph depicts the relative ITD values $\tau/\tau_{\max}$, bounded to the interval $(-1, 1)$. In this experiment, the microphone spacing $2k$ was 17 cm.

Assuming that the microphone spacing $2k = 17\,\text{cm}$, speed of sound $c = 343\,\text{m/s}$, $f_R = 501\,\text{Hz}$ and $f_L = 500\,\text{Hz}$, the ITD derivative equals

$$\frac{\dot{\tau}}{\tau_{\max}} = \frac{2}{0.17/c}\frac{1}{1001} \approx 4\,\text{s}^{-1}. \tag{62}$$

Saberi[37] mentions that in a complex tone, for a higher frequency component to maintain the same rate of change in interaural delay as a lower component, the difference in frequency at the contralateral ear must be proportionately larger. The factor of proportionality can easily be determined by rearranging Eq. (34),

$$f_R = \frac{2+\dot{\tau}}{2-\dot{\tau}} f_L. \tag{63}$$

In the example of Eq. (62), the factor of proportionality in Eq. (63) is 1.002. Therefore, if $f_L = 1000\,\text{Hz}$, the contralateral frequency must be $f_R = 1002\,\text{Hz}$. The second curve in Fig. 13 shows that, in the case of these frequencies, the rate of change in ITDs is the same as for 500 and 501 Hz. However, the appreciation of laterality changes, because at 1 kHz, the signals are out of phase by 180° after 0.25 s. At 500 Hz, this phase shift is reached only after 0.5 s.

In conclusion, the binaural Doppler equation shows that binaural pitch differences are directly related to dynamic sound localization.

IV. DISCUSSION

From the epistemological point of view this work clearly follows a deductive method by attempting to develop a model for binaural localization of sound that does not primarily rely on empirical data, but on a mathematical construct. The presented model requires a few scaffolding conditions. The first requirement consists of the (over)simplification of the binaural system to a couple of acoustic receivers in the free field. Such a system does not exist in nature, nor can it be made artificially. Even with the best design, real microphones are not reduced to infinitesimal spots. Acoustic signals are never generated on such a spot either. Sound waves represent complex phenomena that are affected by diffraction in the presence of a head causing important bias to the fundamental formula relating ITDs to the slope of the cone of confusion [Eq. (2)]. For that reason Lambert[11] and Duda *et al.*[38] adapted this equation to spherical or ellipsoidal heads. Regarding such alternative conditions of the initial geometry, further studies could examine the present model development from the aspect of relevant variants on Eq. (2).

Another constraint for this model is the far field assumption, which allows a few essential simplifications in the equation development. This somehow vague condition has been made precise with two key sound distances $4k$ and $12k$ yielding deviations between the slope of the cone of confusion and the true sound direction less than 1° and 0.1°, respectively.[2] Further investigation of the near field could start from the work by Shinn-Cunningham *et al.*,[39] who suggest considering a torus of confusion for sound sources within the listeners reach.

An additional implicit assumption in the present paper is that acoustic signals received by the microphones are identical in shape and spectrum, except for the modulation effects due to the Doppler shift and the time delays caused by path differences of arrival. In all of the practical experiments made during the development of this work, it appeared that in ordinary room environments this condition could not be reliably controlled. In fact, disturbances such as reverberation interference and standing waves do not necessarily display in the same manner in each of the binaurally received signals, mostly because the microphones occupy different locations in space, introducing unpredictable errors in ITD measurements. Hartmann[40] reports that the quality of sound localization in rooms depends on many different parameters such as direct-signal-to-reverberant-noise ratio, room geometry, and presence of strong attack transients. This explains the choices of an empty cinema theater as the room environment, and a music playing radio as the sound source used in the experiments of this work. Fortunately, the application of the normalized cross-correlation function, as the chosen method for the determination of ITDs, delivers a measure for the quality of every single ITD estimation. The maximum of this function defines the degree of coherence between both audio signals[41,42] and may therefore be used as a weight in an adaptive digital filter, as mentioned in Sec. II C. Note that the time-domain method deployed in our experiments did not allow the separation of multiple sound sources.

In order for Eq. (7) to yield a unique solution for β at any moment in time, the value of the ITD τ may not be zero, which is equivalent to saying that the sound source may not be located in the median plane (O, y, z). Also, angular speed ω_β must be non-zero as well. However, according to Eqs. (1) and (2), the trivial case where $\tau = 0$, yields $\beta = \pm\pi/2$. Moreover, because Eq. (4) always holds, the sign is resolved in this equation, if the binaural system is moving with a known direction of rotation.

The analysis of error propagation in Sec. II B unveiled that in the described model, uncertainty in azimuth dramatically grows as the horizontal rotation speed slows to zero. This does not necessarily apply to human hearing, where different sources of information, such as ITD, ILD, head-related transfer function (HRTF), visual and motional cues, come into play. For instance, Hirahara *et al.*[15] showed that sound localization accuracy in the horizontal or median plane was better with the head rotating (even for slow rotations) than when the head was stationary. Further experimental research could focus on the estimation of the error bounds that are related to human processing of ITD and its derivative.

If the state of the binaural system is defined by its orientation in space with respect to the location of the static sound source, there is confusion about the initial definition of the binaural coordinate system, because only the x axis is unequivocal. However, as soon as a system rotation is performed, it stands to reason that the plane of rotation should be chosen as the "horizontal" plane. In this case, the system

is observable for both azimuth and elevation except for the disambiguation about the upper or lower hemisphere.

System observability is seriously affected in the case of motion of the sound source, which was explicitly excluded in Sec. II, and reintroduced in Sec. III of this paper. The binaural system cannot draw inferences about the instantaneous speed of the sound source, and as a consequence, Eq. (7) is underdetermined. The reference paper by Kneip and Baumann[2] mentions the possibility of completely solving the localization problem through combined motion sequences. In the case of source motion, such a sequence could override the system underdetermination. In fact, if the binaural system is controlled with the source centering algorithm displayed in Table I, it will always keep the moving sound source in the median plane by tracking it. Without any reference to the "world" coordinate system this means that the source will appear immobile in the binaural reference frame. Any additional deliberate rotation of the system with known angular speed will therefore solve the localization problem within the limits explained in this paper.

The binaural Doppler equation developed in Sec. III identifies weak pitch differences between the left and right signals with the variation of ITDs. In other words, a direct relationship has been established between spectral displacements in the binaurally captured signals and sound localization due to system motion. Yet, it can be concluded from the definition of $\dot{\tau}$ [Eq. (3)] that, at any moment in time, there is a unique value of the ITD derivative for a known angular speed and a specific sound direction in each lateral hemisphere. Therefore, two or more sound sources that are located on different cones of confusion will produce different patterns of the binaural Doppler shift. We predict that it might be possible to recognize these patterns in spectral analyses of the auditory scene, which therefore could contribute to the separation of multiple sound sources. A good starting point for further research might be the work by Roman and Wang.[43]

A final note must be addressed about the binaural Doppler shift. Despite the promising way of determining ITD derivatives indirectly via pitch differences, the authors did not follow this approach in their experimental part, because they considered the field of pitch detection beyond the scope of the present work.

V. SUMMARY

In the first part, this paper developed the mathematical model for dynamic sound localization in space in the case of rotations of the binaural baseline based on the knowledge of the rotation speed and the differentiation of ITDs with respect to time, showing that the sound direction in the plane of rotation is exactly and instantaneously determined, whereas its orthogonal direction is only yielded in magnitude without the specification of its sign. The paper also evaluated the error propagation through the established equations, clarifying that uncertainties in the plane of rotation are bounded and mostly depend on the sound direction in the orthogonal plane and the rotation speed. By contrast, uncertainties in the orthogonal plane were proven to grow excessively for sound sources located either in that plane or in the plane of rotation. Finally, the work presented experiments, in order to examine the practical value of the model for embedded real-time applications. The experiments did not include comparative studies with other methods, but focused on the solution of one major issue related to the approach, namely, the synchronization of ITD values and the values of the first ITD derivative. The second part of the paper developed the binaural Doppler equation stating that the first ITD derivative is equivalent to the Doppler effect appearing between the two binaural receivers. This equation was used to describe the complementary frequency modulation taking place at each receiver and the binaural beat phenomenon in the case of small frequency differences.

ACKNOWLEDGMENTS

The authors thank Stephanie Malek and Timothy P. Martin, Department of Mechanical Engineering, Tufts University Medford, MA.

NOMENCLATURE

$R(k,0,0)$	Position of the right microphone, $k>0$
$L(-k,0,0)$	Position of the left microphone
$2k$	Spacing between the two microphones
c	Sound velocity ($\approx$343 m/s in air)
$S(x,y,z)$	Sound-source location
$S_{xy}(x,y,0)$	Orthogonal projection of S on the horizontal plane
$S_x(x,0,0)$	Orthogonal projection of S on the x axis
τ	Interaural time delay (ITD)
$\tau_{max}=2k/c$	Maximal detectable ITD
$\dot{\tau}$	Derivative of ITD with respect to time
d_R, d_L, d	Distance to sound source equal to $\|\vec{RS}\|, \|\vec{LS}\|, \|\vec{OS}\|$
$d_L-d_R=c\tau$	Path difference of arrival ($c=$ sound velocity)
v_{SR}, v_{SL}	Speed of the sound source toward the microphones, equal to the derivatives $\dot{d}_R$, $\dot{d}_L$
v_R, v_L	Speed of the microphones toward the sound source, equal to $-v_{SR}$, $-v_{SL}$
s_A	Acoustic signal at the source
s_R, s_L	Acoustic signals at the microphones
f_A, ω_A, λ_A	Frequency, angular frequency and wavelength of the acoustic signal at its source
$f_R, \lambda_R, f_L, \lambda_L$	Frequencies and wavelengths of the acoustic signal as detected on each microphone
$\varphi \in [0,\pi]$	True sound direction in the plane (R, L, S), referred to the origin O
$\phi \in [0,\pi]$	Slope of the cone of confusion, $\phi \simeq \varphi$ for sound sources in the far field
ϕ_R, ϕ_L	Sound directions referred to the microphones R, L
$\omega_\phi=\dot{\phi}$	Variation of sound direction in the plane (R, L, S)

$\psi \in [-\pi/2, \pi/2]$ Elevation, vertical angle between the horizontal plane and the sound-source vector

$\beta \in [-\pi, \pi]$ Horizontal angle (=azimuth), referenced to the x axis

$\omega_\beta = \dot{\beta}$ Variation of azimuth over time, i.e., rotation speed about the z axis

θ_z Rotation angle about the z axis in "world" referential frame

APPENDIX A: EQUATIONS (8) AND (9)

$$\begin{aligned}
&\int_{t_0}^{t_1} \frac{\dot{\tau}(t)}{\tau(t)} \mathrm{d}t = \int_{t_0}^{t_1} -\dot{\beta}(t) \tan\beta(t) \mathrm{d}t \\
&\quad \Longleftrightarrow \left[\ln|\tau(t)|\right]_{t_0}^{t_1} = \left[\ln|\cos\beta(t)|\right]_{t_0}^{t_1} \\
&\quad \Longleftrightarrow \ln\left|\frac{\tau_1}{\tau_0}\right| = \ln\left|\frac{\cos\beta_1}{\cos\beta_0}\right| \\
&\quad \Longleftrightarrow \frac{\tau_1}{\tau_0} = \frac{\cos\beta_1}{\cos\beta_0} \text{ or } \frac{\tau_0}{\tau_1} = \frac{\cos\beta_0}{\cos\beta_1}.
\end{aligned}$$

Because $\beta_0 = \beta_1 + \theta_z$ [cf. Eq. (6)], this changes to (provided $\tau_0 \neq 0$, $\tau_1 \neq 0$ and $\theta_z \neq 0$)

$$\begin{aligned}
\frac{\tau_0}{\tau_1} &= \frac{\cos\beta_1 \cos\theta_z - \sin\beta_1 \sin\theta_z}{\cos\beta_1} \\
&= \cos\theta_z - \tan\beta_1 \sin\theta_z \\
&\Longleftrightarrow \frac{\tau_0}{\sin\theta_z \tau_1} = \cot\theta_z - \tan\beta_1 \\
&\Longleftrightarrow \tan\beta_1 = \cot\theta_z - \frac{\tau_0}{\sin\theta_z \tau_1}.
\end{aligned}$$

APPENDIX B: EQUATION (49)

$$\begin{aligned}
\dot{\phi} &= \frac{x^2}{x^2+y^2} \frac{x\dot{y} - \dot{x}y}{x^2} \\
&= \frac{xv_y - yv_x}{d^2} \\
&= \frac{v_y \cos\phi - v_x \sin\phi}{d}.
\end{aligned}$$

APPENDIX C: EQUATION (53)

$$\begin{aligned}
v_R &= -v_{SR} = -\dot{d}_R \\
&= -\frac{\dot{A} - \left(\dot{B}\cos\phi + B\dot{\cos}\phi\right)}{2d_R} \\
&= -\frac{2d\dot{d} - \left(2k\dot{d}\cos\phi - 2kd\dot{\phi}\sin\phi\right)}{2d_R} \\
&= -d\left(v_x\cos\phi + v_y\sin\phi\right) - k\left(v_x\cos^2\phi + v_y\sin\phi\cos\phi\right) \\
&\quad + k\left(v_y\sin\phi\cos\phi - v_x\sin^2\phi\right)/d_R \\
&= -\frac{d\left(v_x\cos\phi + v_y\sin\phi\right) - kv_x}{d_R} = -\frac{d\dot{d} - kv_x}{d_R}.
\end{aligned}$$

APPENDIX D: EQUATION (55)

$$\begin{aligned}
v_R + v_L &= -\frac{d_L\left(d\dot{d} - kv_x\right) + d_R\left(d\dot{d} + kv_x\right)}{d_L d_R} \\
&= -\frac{(d_L + d_R)d\dot{d} - (d_L - d_R)kv_x}{d^2 - k^2} \\
&\simeq -\frac{2d^2\left(v_x\cos\phi + v_y\sin\phi\right) - 2k^2 v_x\cos\phi}{d^2 - k^2} \\
&= -2\left(v_x\cos\phi + \frac{d^2}{d^2-k^2} v_y\sin\phi\right) \\
&= -2\left(v_x\cos\phi + \frac{1}{1-k^2/d^2} v_y\sin\phi\right) \\
&\simeq -2\left(v_x\cos\phi + v_y\sin\phi\right).
\end{aligned}$$

APPENDIX E: LEAST-SQUARE LOW-PASS FILTER

Equation (16) in Sec. II C is derived according to Schüßler.[26] In contrast to the method used by this author, the low-pass filter is developed here to the first order only, which is why ordinary algebraic equations appear more appropriate than the original vector notation.

Assume that $s(t)$ describes a noisy signal in function of time. A least-square filter for this function can be designed as a discrete polynomial approximation $p[j]$ that is evaluated at M values of the input signal $s(t)$, which have been taken at equal time steps Δt with respect to the current time $t_{M-1} = (M-1)\Delta t$. The equation of such a first-order polynomial may be written as

$$s_j = s(t_j) \approx p[j] = a_0 + a_1 t_j. \tag{E1}$$

Ultimately, if $\Delta t \to 0$, the segment defined by $p[j]$ can be considered as a superposable model for the signal curve within the same interval. However, because this model is not correct for finite values of Δt, the best linear fit can be obtained by minimizing the sum of the squared residuals Σ_r,

$$\Sigma_r = \sum_{j=0}^{M-1} \left[s_j - (a_0 + a_1 t_j)\right]^2. \tag{E2}$$

The sum Σ_r is unknown. However, it can be considered a function of both values a_0 and a_1, and its minimum can be found, if both partial derivatives are zeroed:

$$\begin{cases}
\dfrac{\partial\Sigma_r}{\partial a_0} = -2\displaystyle\sum_{j=0}^{M-1}\left[s_j - (a_0 + a_1 t_j)\right] = 0 \\
\dfrac{\partial\Sigma_r}{\partial a_1} = -2\displaystyle\sum_{j=0}^{M-1} t_j\left[s_j - (a_0 + a_1 t_j)\right] = 0.
\end{cases} \tag{E3}$$

Gauss's method of the least-squares yields a_0 and a_1 as the solution of the system of equations that follows from Eq. (E3),[44]

$$\begin{cases} Ma_0 + (\sum t_j)a_1 = \sum s_j \\ \left(\sum t_j\right)a_0 + \left(\sum t_j^2\right)a_1 = \sum t_j s_j. \end{cases} \tag{E4}$$

Using the summation terms $c_1 = \sum t_j$, $c_2 = \sum t_j^2$, and $c_3 = Mc_2 - c_1^2$, the solutions for a_0, a_1 may be written as

$$\begin{aligned} a_0 &= \frac{c_2 \sum s_j - c_1 \sum t_j s_j}{c_3}, \\ a_1 &= \frac{M \sum t_j s_j - c_1 \sum s_j}{c_3}. \end{aligned} \tag{E5}$$

Under the assumption of constant Δt, the terms of Eq. (E5) may be rewritten,

$$\begin{aligned} c_1 &= \sum_{j=0}^{M-1} t_j = \sum_{j=0}^{M-1} j\Delta t = \Delta t \sum_{j=0}^{M-1} j \stackrel{!}{=} \Delta t c_1', \\ c_2 &= \sum_{j=0}^{M-1} t_j^2 = \Delta t^2 \sum_{j=0}^{M-1} j^2 \stackrel{!}{=} \Delta t^2 c_2', \\ c_3 &= Mc_2 - c_1^2 = \Delta t^2 (Mc_2' - c_1'^2) \stackrel{!}{=} \Delta t^2 c_3', \\ \sum t_j s_j &= \Delta t \sum j s_j, \end{aligned} \tag{E6}$$

leading to the equations

$$\begin{aligned} a_0 &= \frac{c_2' \sum s_j - c_1' \sum j s_j}{c_3'}, \\ a_1 &= \frac{M \sum j s_j - c_1' \sum s_j}{\Delta t c_3'} \end{aligned} \tag{E7}$$

The summations of consecutive integers and their squares are standard operations of sequences that can be easily pre-calculated.

$$\begin{aligned} c_1' &= \sum_{j=0}^{M-1} j = M(M-1)/2, \\ c_2' &= \sum_{j=0}^{M-1} j^2 = M(M-1)(2M-1)/6, \end{aligned} \tag{E8}$$

which is why

$$\begin{aligned} a_0 &= \frac{2(2M-1)\sum s_j - 6\sum j s_j}{M(M+1)}, \\ a_1 &= \frac{12\sum j s_j - 6(M-1)\sum s_j}{\Delta t \cdot M(M^2-1)}. \end{aligned} \tag{E9}$$

Because the polynomial $p[j]$ should be evaluated at t_{M-1}, in order to yield the most actual estimation,

$$p[M-1] = a_0 + a_1(M-1)\Delta t = \frac{\sum_{j=0}^{M-1} \big(6j - 2(M-2)\big) s_j}{M(M+1)}. \tag{E10}$$

This low-pass filtering technique can be developed for higher orders using Faulhaber's sums of powers formula (cf. Knuth[45]).

[1] K. Nixdorff, *Mathematische Methoden der Schallortung in der Atmosphäre* (*Mathematical Methods of Sound Localization in the Atmosphere*) (Vieweg-Verlag, Braunschweig, Germany, 1977), pp. 5–11.
[2] L. Kneip and C. Baumann, "Binaural model for artificial spatial sound localization based on interaural time delays and movements of the interaural axis," J. Acoust. Soc. Am. **124**, 3108–3119 (2008).
[3] C. F. Altmann, E. Wilczek, and J. Kaiser, "Processing of auditory location changes after horizontal head rotation," J. Neurosci. **29**, 13074–13078 (2009).
[4] J. L. van Soest, "Richtungshooren bij sinusvormige geluidstrillingen" ("Directional hearing of sinusoidal waves"), Physica (The Hague) **9**, 271–282 (1929), available at http://opac.nebis.ch.
[5] G. L. Reid and E. Milios, "Active stereo sound localization," J. Acoust. Soc. Am. **113**, 185–193 (2003).
[6] H. Wallach, "The role of head movements and vestibular and visual cues in sound localization," J. Exp. Psychol. **27**, 339–368 (1940).
[7] S. Perret and W. Noble, "The effect of head rotations on vertical sound localization," J. Acoust. Soc. Am. **102**, 2325–2332 (1997).
[8] F. Whightman and D. J. Kistler, "Resolution of front-back ambiguity in spatial hearing by listener and source movement," J. Acoust. Soc. Am. **105**, 2841–2853 (1999).
[9] W. R. Thurlow, J. W. Mangels, and P. S. Runge, "Head movements during sound localization," J. Acoust. Soc. Am. **42**, 489–493 (1967).
[10] P. A. Hill, P. A. Nelson, O. Kirkeby, and H. Hamada, "Resolution of front-back confusion in virtual acoustic imaging systems," J. Acoust. Soc. Am. **108**, 2901–2910 (2000).
[11] R. M. Lambert, "Dynamic theory of sound-source localization," J. Acoust. Soc. Am. **56**, 165–171 (1974).
[12] D. Rao and B. Xie, "Head rotation and sound image localization in the median plane," Chin. Sci. Bull. **50**, 412–416 (2005).
[13] E. A. Macpherson, "Cue weighting and vestibular mediation of temporal dynamics in sound localization via head rotation," Proc. Meet. Acoust. **19**, 050131 (2013).
[14] D. Morikawa, Y. Toyoda, and T. Hirahara, "Head movement during horizontal and median sound localization experiments in which head-rotation is allowed," J. Acoust. Soc. Am. **133**, 3510–3510 (2013).
[15] T. Hirahara, D. Yoshisaki, and D. Morikawa, "Impact of dynamic binaural signal associated with listener's voluntary movement in auditory spatial perception," J. Acoust. Soc. Am. **133**, 3459–3459 (2013).
[16] J. Blauert, *Spatial Hearing*, revised edition (MIT Press, Cambridge, MA, 1983), pp. 178–202.
[17] E. M. von Hornbostel and M. Wertheimer, "Über die Wahrnehmung der Schallrichtung" ("On the perception of the direction of sound"), Sitzungsber. K. Preuss. Akad. Wiss. **20**, 388–396 (1920), available at https://ia802607.us.archive.org/33/items/sitzungsberichte1920preu/sitzungsberichte1920preu.pdf.
[18] S. B. Andersson, A. A. Handzel, V. Shah, and P. S. Krishnaprasad, "Robot phonotaxis with dynamic sound-source localization," in *Proceedings of the 2004 IEEE International Conference on Robotics and Automation* (2004), Vol. 5, pp. 4833–4838.
[19] P. X. Zhang and W. M. Hartmann, "Lateralization of sine tones-interaural time vs phase," J. Acoust. Soc. Am. **120**, 3471–3474 (2006).
[20] A. Portello, P. Danès, and S. Argentieri, "Acoustic models and Kalman filtering strategies for active binaural sound localization," in *Proceedings of the 2011 IEEE International Conference on Intelligent Robots and Systems* (2011), pp. 137–142.
[21] M. Grabe, *Measurement Uncertainties in Science and Technology* (Springer, New York, 2005), pp. 70–90.
[22] B. M. Sayers and E. C. Cherry, "Mechanism of binaural fusion in the hearing of speech," J. Acoust. Soc. Am. **29**, 973–987 (1957).
[23] U. Meyer-Baese, *Digital Signal Processing With Field Programmable Gate Arrays*, 3rd ed. (Springer, New York, 2007), 518 pp.
[24] D. W. Grantham and F. L. Wightman, "Detectability of a pulsed tone in the presence of a masker with time-varying interaural correlation," J. Acoust. Soc. Am. **65**, 1509–1517 (1979).
[25] D. R. Kincaid and E. W. Cheney, *Numerical Analysis, Mathematics of Scientific Computing*, 3rd ed. (American Mathematical Society, Providence, RI, 2002), pp. 28–41.
[26] H. W. Schüßler, *Digitale Signalverarbeitung 2* (*Digital Signal Processing 2*) (Springer, Heidelberg, Germany, 2010), pp. 376–381.

[27]R. L. Jenison, "On acoustic information for motion," Ecol. Psychol. **9**, 131–151 (1997).
[28]R. Müller and H.-U. Schnitzler, "Acoustic flow perception in cf-bats: Properties of the available cues," J. Acoust. Soc. Am. **105**, 2958–2966 (1999).
[29]M. F. Neelon and R. L. Jenison, "The effect of trajectory on the auditory motion aftereffect," Hearing Res. **180**, 57–66 (2003).
[30]Y. Iwaya, M. Toyoda, and Y. Suzuki, "A new rendering method of moving sound with the Doppler effect," in *11th International Conference on Auditory Display*, Limerick, Ireland (2005) pp. 253–255.
[31]M. Kumon and S. Uozumi, "Binaural localization for a mobile sound source," J. Biomech. Sci. Eng. **6**, 26–30 (2011).
[32]W. M. Hartmann, *Signals, Sound and Sensation* (Springer, New York, 1998), 430 pp.
[33]G. Plenge, "On the differences between localization and lateralization," J. Acoust. Soc. Am. **56**, 944–951 (1974).
[34]Lord Rayleigh, "On our perception of sound direction," Philos. Mag. **13**, 214–232 (1907).
[35]J. C. R. Licklider, J. C. Webster, and J. M. Hedlun, "On the frequency limits of binaural beats," J. Acoust. Soc. Am. **22**, 468–473 (1950).
[36]M. A. Akeroyd, "A binaural beat constructed from a noise," J. Acoust. Soc. Am. **128**, 3301–3304 (2010).
[37]K. Saberi, "Fast Fourier-based dsp algorithm for auditory motion experiment," Behav. Res. Methods Instrum. Comput. **36**, 585–589 (2004).
[38]R. O. Duda, C. Avendano, and V. R. Algazi, "An adaptable ellipsoidal head model for the interaural time difference," in *Proceedings of the IEEE International Conference of the Acoustics, Speech, and Signal Processing* (1999), pp. 965–968.
[39]B. G. Shinn-Cunningham, S. Santarelli, and N. Kopco, "Tori of confusion: Binaural localization cues for sources within reach of a listener," J. Acoust. Soc. Am. **107**, 1627–1636 (2000).
[40]W. M. Hartmann, "Localization of sound in rooms," J. Acoust. Soc. Am. **74**, 1380–1391 (1983).
[41]B. Rakerd and W. M. Hartmann, "Localization of sound in rooms. V. binaural coherence and human sensitivity to interaural time differences in noise," J. Acoust. Soc. Am. **128**, 3052–3063 (2010).
[42]C. Faller and J. Merimaa, "Source localization in complex listening situations: Selection of binaural cues based on interaural coherence," J. Acoust. Soc. Am. **116**, 3075–3089 (2004).
[43]N. Roman and D. Wang, "Binaural tracking of multiple moving sources," Trans. Audio Speech Lang. Proc. **16**, 728–739 (2008).
[44]J. O. Rawling, S. G. Pantula, and D. A. Dickey, *Applied Regression Analysis: A Research Tool*, 2nd ed. (Springer, New York, 1998), pp. 3–7.
[45]D. E. Knuth, "Johann Faulhaber and sums of powers," Math. Comp. **61**, 277–294 (1993).