Audio Engineering Society

Convention Express Paper 348

Presented at the AES 158th Convention
2025 May 22–24, Warsaw, Poland

This Express Paper was selected on the basis of a submitted synopsis that has been peer-reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This Express Paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. This paper is available in the AES E-Library (http://www.aes.org/e-lib), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Instantaneous low-frequency energetic analysis for detection of standing waves

Mădălina Năstasă[1,2], Ville Pulkki[1], and Aki Mäkivirta[2]

[1]*Department of Information and Communications Engineering, Aalto University, Espoo, Finland*
[2]*Genelec Oy, Iisalmi, Finland*

Correspondence should be addressed to Mădălina Năstasă (amm.nastasa@gmail.com)

ABSTRACT

The ability to hear direction of reproduced low-frequency sound is critically influenced by standing-wave phenomena in enclosed spaces, where standing waves can become the dominant directional cues and overpower the cues present in reproduced audio. Conventional measurement methods struggle to identify room modes. The aim of this study is to introduce a single-point instantaneous energetic analysis able to identify standing waves using a combination of energy-estimation parameters. We propose a standing-wave estimator Γ based on two diffuseness metrics. The estimator is evaluated through numerical 2D simulations covering a wide range of room locations, including both node and antinode positions of a room mode standing wave resonance. Γ can identify mode resonant frequencies and indicate the directions of mode resonances. Combined with magnitude response and temporal decay analyses, Γ has the potential to provide a more comprehensive understanding of low-frequency source behaviour in enclosed spaces.

1 Introduction

The acoustical behaviour of low-frequency sound sources in enclosed spaces is a matter deserving of attention due to its critical role in both optimising sound reproduction uniformity and enhancing the ability to provide accurate auditory directional cues at these frequencies.

The directional perception of low frequencies has been a matter long debated [1, 2, 3, 4, 5, 6], where the methodology of testing and the results obtained have drastically varied. Research indicating that low frequencies may lack in providing directional cues [7, 8, 9, 10, 11, 12] have typically been conducted in small, living-room-like listening environments [8, 9, 10]. These environments often involve conditions where the position of sound sources generates strong modal behaviour. Several authors have observed that localisation ability might be hindered by the presence of room modes, which could adversely

affect the binaural cues received by the listener. In addition, recent research [13] conducted in an anechoic chamber with a small amount of modal resonant behaviour highlighted that the localisation ability in this frequency region is degraded as a result of conflicting auditory cues offered to the auditory system by the presence of standing waves in the listening environment.

Standing waves are a phenomenon ever-present in the sound field of the low-frequency spectrum, which are especially difficult to measure and quantify due to their temporal and spatial characteristics. Their presence can severely confound the perceived direction of arrival of sound, making it difficult to accurately determine the origin of sound waves. This phenomenon is especially problematic in the presence of a node within a destructive interference field, which can greatly diminish the ability to localise sound sources. Previous research conducted by the authors [14] suggests that there is a relationship between the presence of standing waves and the low-frequency localisation ability, where elevated modal activity significantly affected the localisation of very low-frequency sound sources and contributed to the confounding of directional cues in this frequency range. Thus, determining the presence and characteristics of standing waves within the listening environment is crucial for low-frequency reproduction if localisation ability is desired.

In this work, a novel method to detect the presence of standing waves is derived by means of instantaneous energetic analysis of the sound field. Two estimators of diffuseness, namely energy density and temporal variation, are utilised in standing wave fields by making use of its sound field properties and phenomena.

2 Standing waves

Room modes are resonances that arise at the natural resonant frequencies of a room when the room is excited by acoustic sources.

Consider two parallel walls. The axial standing wave between the walls (Fig. 1) represents a simple room mode. Here, sound reflected from the two walls interferes to generate stationary pressure pattern between the walls. Axial modes occur at integer multiples of the half-wavelength of sound fitting between the walls, at frequencies

$$f_n = nc/2d, \tag{1}$$

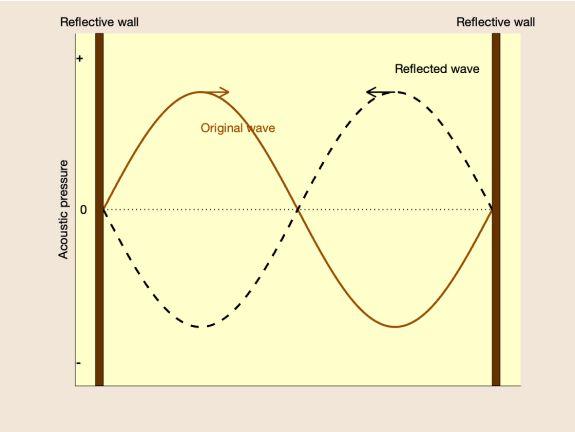


Fig. 1: Generation of a standing wave between parallel reflective surfaces, figure reproduced from [15]

where n= 1,2,3,. . . , c is the speed of sound, and d is the distance between the walls. In enclosed spaces, room modes can occur in more complex ways. For example, in a shoe-box room shape, modes between two-wall corners are called tangential modes, and modes between three-wall corners are called oblique modes.

The pressure pattern in a mode resonance shows constructive (maximum pressure point or antinode) and destructive interference (minimum pressure point or node). Constructive interference amplifies sound level, while destructive interference reduces sound level. This results in a location-dependent pattern of higher and lower sound pressure levels within the room.

The ITD and ILD cues in a room showing significant mode resonances can yield auditory images unrelated to the sound source reproduced in the room. Mode resonances store energy efficiently, making the sound level differences at modes higher and the decay of sound level slower at mode resonant frequencies, changing the perception of sound colour. Reproduction of recorded direction of sound can be overpowered by the direction of sound in mode resonances, rendering the recorded directional information inaudible at low frequencies.

3 Energetic analysis

This section presents the energetic analysis used to characterise sound fields based on sound intensity and diffuseness methods. Furthermore, emphasis is placed on the behaviour of sound intensity components in the low-frequency spectrum.

3.1 Sound Intensity

Sound intensity quantifies the flow of acoustic energy within a sound field [16]. More specifically, it is a vector quantity having a direction and magnitude representing the net flow of sound energy. The instantaneous sound intensity, at time t, is defined as the product of sound pressure p and particle velocity u [16, 17],

$$I(t) = p(t)u(t) \tag{2}$$

The opposite direction of the intensity vector can serve as the estimation for the direction of arrival of sound.

In a monochromatic (single-frequency) stationary sound field, the active intensity is determined by the phase relationship between the time-harmonic sound pressure $p(t)$ and the particle-velocity vector $u(t)$. This arises because active intensity is defined as the time-average of the instantaneous product $p(t)u(t) = ||\langle I(t)\rangle||$, and any phase shift between these quantities alters the sign and direction of that product over a cycle. The active component of instantaneous intensity is proportional to the spatial gradient of phase [18]. By contrast, if there is a phase difference between $p(t)$ and any directional component of $u(t)$, this will result in an oscillation of the direction of the intensity vector [18].

For an ideal travelling plane wave, the particle velocity $u(t)$ and the pressure $p(t)$ are directly proportional, their magnitude oscillating either exactly in phase or out of phase. Consequently, $p(t)u(t)$ retains a constant sign and direction throughout the cycle, indicating that all energy propagates in the same direction.

For an ideal standing-wave field—formed by two equal-amplitude plane waves propagating in opposite directions—the acoustic pressure and particle velocity are everywhere 90°out of phase. Consequently, the active intensity vanishes throughout the field, indicating no net energy flow. By contrast, the reactive intensity remains non-zero and represents the oscillatory exchange of energy between the pressure and the particle velocity components at each point. Unlike active intensity, reactive intensity describes energy that is temporarily stored and returned each cycle, characterising purely reactive acoustic behaviour without any net propagation. Therefore, the reactive component of instantaneous intensity depends on the spatial gradient of the amplitude of the sound pressure [18].

3.2 Diffuseness estimators

A diffuse sound field is commonly defined as one in which the instantaneous flow of sound energy is equally probable in all directions, and density of sound energy is uniform throughout the field [17, 19]. Then, there is no net transport of energy, and average intensity $\langle I\rangle$ across the field is zero.

In contrast, a plane wave propagating in free field represents the ideal opposite of a diffuse sound field. A standing wave can be modelled as two plane waves propagating in opposite directions. Generally then, a low-frequency sound field within a room can be represented as a combination of standing waves and a diffuse field component.

3.2.1 Energy density-based estimator of diffuseness

In energetic analysis, diffuseness ψ is traditionally estimated as the ratio between the magnitude of the net sound energy flow and the total sound energy density [18],

$$\psi = 1 - \frac{||\langle I\rangle||}{\langle cE\rangle}, \tag{3}$$

where $\langle\cdot\rangle$ denotes time averaging, and $||\cdot||$ is the norm of the vector. E is a scalar quantity that describes the energy density at a certain point of the sound field,

$$E(t) = \frac{1}{2}\rho_0(Z_0^{-2}p(t)^2 + u(t)^2), \tag{4}$$

where ρ_0 is the mean density of air, and $Z_0 = \rho_0 c$ is the acoustic impedance of air, c being the speed of sound.

In free field, the magnitude of the active intensity vector of a plane wave equals the product of energy density and speed of sound, resulting in diffuseness ψ value of zero in Eq. 3. In contrast, when multiple incoherent plane waves arrive from different directions, the magnitude of the intensity vector becomes, on average, smaller than the product of the energy density and the speed of sound. As a result, the analysed energy-based diffuseness estimator increases towards unity.

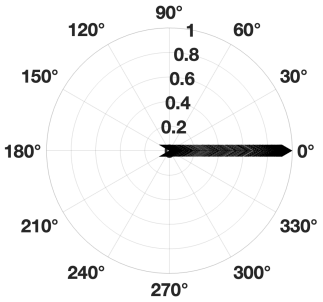
3.2.2 Temporal variation estimator of diffuseness

The temporal variation estimator $\hat{\psi}$, introduced by Ahonen and Pulkki [20], utilises the temporal variation in

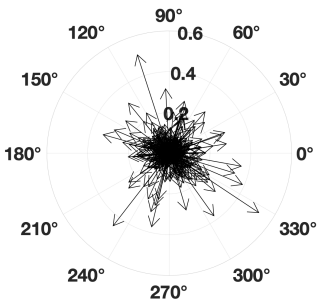
the intensity vector. The estimator is defined by dividing the length of averaged intensity vectors with the averaged length of the same vectors as

$$\hat{\psi} = \sqrt{1 - \frac{||\langle I(t)\rangle||}{\langle||I(t)||\rangle}} \tag{5}$$

The motivation for this estimator is illustrated in Fig. 2. In the case of a purely diffuse field, $||\langle I(t)\rangle|| = 0$, because the directional variation of the vectors averages to zero, resulting in temporal variation $\hat{\psi} = 1$. Conversely, for a plane wave, $||\langle I(t)\rangle|| = \langle||I(t)||\rangle$ and the temporal variation $\hat{\psi} = 0$.



(a) Plane wave field



(b) Diffuse field

Fig. 2: Simulated intensity vectors estimated over 10 ms at a certain location for (a) a plane plane wave propagating in direction 0° and (b) diffuse field, synthesised using the plane-wave model

4 Standing wave detection by means of diffuseness estimators

4.1 Theory

Reactive acoustic fields are characterised by local oscillations of energy—such as those that occur adjacent to boundaries or within resonant cavities—without any net transport of energy. In this context, local refers to oscillations of energy that are confined to specific spatial regions, without contributing to a net transport of energy across the medium. In a pure standing-wave field, pressure and particle-velocity fluctuations are strong, yet the time-averaged active intensity is identically zero.

The diffuseness estimator is not able to differentiate diffuse and modal fields, but indicates for both zero net active intensity, $||\langle I(t)\rangle|| = 0$. Standing waves yield zero net intensity despite their highly ordered, non-diffuse nature. Consequently, when modal overlap is insufficient and standing-wave modes dominate, time-averaged energy density-based estimators ascribe high diffuseness also to reactive, non-random fields.

The temporal variation estimator, however, can be useful in ascribing standing wave fields, recognizing that a standing wave can be considered to comprise equal-amplitude plane waves propagating in opposite directions. We propose using one-fold and two-fold processes followed by expanding, to enable the temporal variation estimator the ability to detect the standing wave field correctly. After the folding procedures, standing waves are detected as plane waves whilst having no effect on correctly detecting nonstationary and diffuse fields, Fig. 3.

Finally, the energy-based estimator can be combined with the modified temporal variation estimator to produce a new estimator capable of differentiating standing waves created by modal resonances.

4.1.1 Two-Fold and Expand Procedure

The process of folding is applied to the instantaneous intensity vectors for the temporal variation estimator, exploiting the known directional symmetry of standing waves. This folding is defined for a two-dimensional (2D) case. While a similar expansion to three dimensions (3D) is theoretically possible, it is not addressed in this work. First, instantaneous intensity vectors are folded into the first Cartesian quadrant,

$$\begin{aligned} I_x(t) &= |I_x(t)| \\ I_y(t) &= |I_y(t)| \end{aligned} \tag{6}$$

Moving back to polar representation, Eq. 7, the data now describes the magnitude and direction of energy flow within just one quadrant,

$$\theta = \arctan(I_y, I_x)$$
$$r = \sqrt{I_x^2 + I_y^2} \tag{7}$$

The one-quadrant data is now expanded to cover the full circle by multiplying the angle by four, Eq. 8. This maps the $[0, \frac{\pi}{2}]$ interval onto the entire $[0, 2\pi]$ range,

$$\theta = \theta \cdot 4 \tag{8}$$

The modified Cartesian intensity components can be obtained as Eq. 9.

$$I_x = r \cdot \cos(\theta)$$
$$I_y = r \cdot \sin(\theta) \tag{9}$$

4.2 One-Fold and Expand Procedure

In this procedure, folding is achieved by mirroring negative values of the first spatial component (I_x) onto the positive domain, implemented by

$$I_x(t) = |I_x(t)| \tag{10}$$

Similarly to the first procedure, next we convert to polar coordinates, Eq. 7, and the angle θ now ranges $[-\frac{\pi}{2}, \frac{\pi}{2}]$. This collapses vectors into the right-hand half-plane $I_x \geq 0$, so all directions now lie on the positive x values.

The folded angle is now expanded onto the entire $[0, 2\pi]$ range,

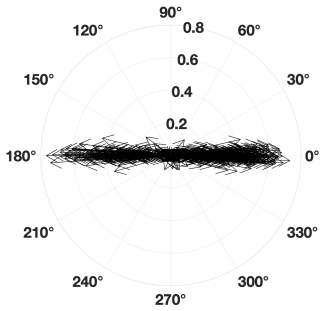
$$\theta = \theta \cdot 2 \tag{11}$$

4.3 Standing wave estimator

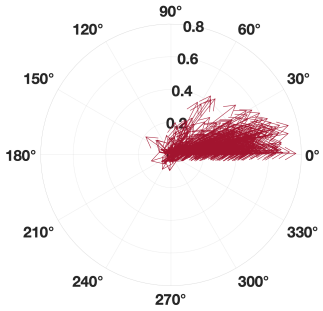
We now define a new estimator Γ that exploits the relationship between the energy-based and modified temporal variation estimators for detecting a standing wave field. The standing wave estimator is defined as

$$\Gamma = \psi(1 - \hat{\psi}), \tag{12}$$

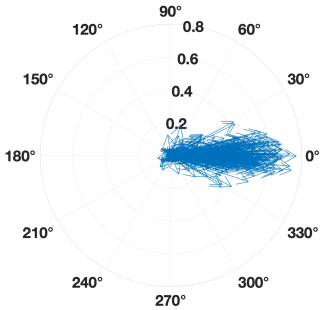
where $\hat{\psi}$ is calculated utilising the instantaneous intensity values that have been processed through the one or two-fold and expand procedure. A high value of Γ indicates a standing wave field, while a low value indicates non-stationary non-reactive sound field.



(a) Original sound field



(b) Result of two-fold expand procedure



(c) Result of one-fold expand procedure

Fig. 3: Illustration of the fold-and-expand procedures applied to a sound field composed of a standing wave aligned at 0 °, combined with a diffuse field 10 dB louder than the standing wave (this level is chosen to improve visualisation clarity)

5 Evaluation

5.1 Methodology

The standing-wave estimator Γ was evaluated using a simulated monochromatic sound field constructed by two plane waves propagating in the x, y- plane, both at frequency f = 80 Hz in air (ρ = 1.2 kg/m^3, c =

343 m/s), yielding angular frequency $\omega = 2\pi f$ and wave-number $k = \frac{\omega}{c}$. The pressure fields are computed using the plane wave solution at location (x,y). The instantaneous pressure is

$$p_i(x,y,t) = A_i \cos(k(x\cos(\theta_i) + y\sin(\theta_i)) - \\ - \omega \cdot t + \phi + \frac{x}{c}) \quad (13)$$

Two plane waves, indexed by i = [1,2], propagate at angles θ_i having amplitudes A_i and initial phases ϕ_i. Here, the term $\frac{x}{c}$ models a small propagation delay. For sake of digital simulation, time-domain signals are sampled at rate F_s = 48 kHz for a duration of one second. To emulate microphone capsule self-noise, Gaussian noise is added to create signal-to-noise of 60 dB in the pressure signals. The average composite pressure is then

$$p(x,y,t) = \frac{p_1 + p_2}{2}. \quad (14)$$

Under the plane-wave approximation, the particle-velocity phasors align with the propagation directions. This yields time-domain component velocities

$$u_{ix}(x,y,t) = p_i(x,y,t)\cos(\theta_i) \\ u_{iy}(x,y,t) = p_i(x,y,t)\sin(\theta_i) \quad (15)$$

which are scaled by the acoustic impedance of air, $Z_0 = \rho_0 c$, to account for the relationship between pressure and particle velocity in the medium.

$$u_x = \frac{u_{1x} + u_{2x}}{Z_0} \\ u_y = \frac{u_{1y} + u_{2y}}{Z_0} \quad (16)$$

The diffuse field was modelled by superposing N = 360 monochromatic plane waves of equal amplitude A, each propagating in randomised directions $\theta_i \in [0, 2\pi]$, with random phase $\phi_i \in [0, 2\pi]$ [21],

$$p_{\text{diffuse}}(t) = \frac{1}{N}\sum_{i=1}^{N} A\cos\Big(k(x\cos\theta_i + y\sin\theta_i) - \\ - \omega t + \phi_i + \tfrac{x}{c}\Big) \quad (17)$$

The diffuse field signal RMS level was scaled to obtain the desired level difference to the standing-wave component.

A standing wave was created by superimposing two plane waves propagating in opposite directions 0 °and 180 °.

The estimator's performance was evaluated while the diffuse field level was varied in the range $[-60, 20]$ dB relative to the standing wave level. Estimators were averaged over one second duration, without applying windowing. Finally, the standing wave estimator Γ was computed in area $x, y \in [-1.5, 1.5]$ m.

5.2 Results

A spatial map of the standing wave estimator Γ values, computed across the two-dimensional domain (Figure 4) shows that the estimator reliably detects the presence of a standing wave 10 dB louder than the diffuse field ($\Delta L = 10$ dB), across the entire domain except at the nodal line $x = 0$ m. At this pressure antinode the acoustic pressure reaches a maximum while the particle velocity approaches zero. Since the standing wave estimator uses instantaneous intensity of both pressure and particle velocity, and as the particle velocity is very low at the antinode, the estimator displays a zero value. The estimator detects only the diffuse field at that location as the standing wave contribution is very low at the antinode. Also the ideal simulation conditions may contribute.

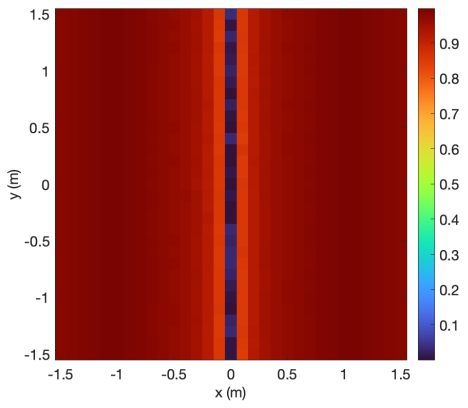


Fig. 4: Spatial map of Γ at ΔL = -10 dB with one-fold and expand procedure in 2D domain ranging from $[-1.5, 1.5]$ m

Comparing the effect of the folding strategies on estimating the Γ (Figure 5) demonstrates that the one-fold procedure is marginally less affected by the diffuse

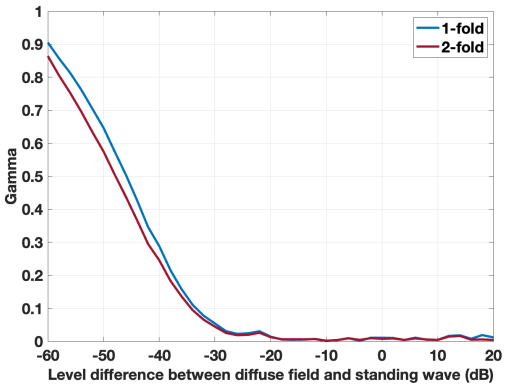


Fig. 5: Γ at (x,y) = (0,0) m as a function of the level difference between the diffuse field and standing wave

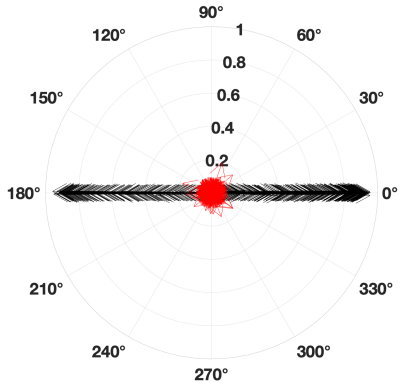


Fig. 6: Instantaneous intensity field at (x, y) = (-1.5, -1.5) m showing the separate contributions of the standing wave (black) and diffuse field (red) components over 10 ms of activity. The fields are presented individually for clarity, without summing them for visualisation purposes

field level, offering marginally better robustness. The value of Γ depends on the level difference between the standing wave and the diffuse field and the measurement position. In practical scenarios, the diffuse field level likely remains low at room mode resonances as standing waves typically decay much slower than the diffuse field at low frequencies.

A compass plot (Figure 6) can visualize low-frequency behaviour by revealing energy flow patterns and modal structures that are not easily revealed in traditional magnitude response and sound decay response measurements. The example shown is visualising the instantaneous intensity vectors over a 10-ms window. Each arrow indicates the magnitude and direction of net energy flow. Black arrows relate to the standing wave component, while red arrows relate to the diffuse field component, having level 10 dB or more below the standing wave component.

While the estimator does not detect the standing wave at the room resonance antinode in the simulations, this must be studied in real rooms. A sufficiently high standing wave relative to the diffuse field does enable the estimator to detect the standing wave even at the antinode, as demonstrated by Fig. 5.

Our study is a two-dimensional model. The performance of the standing wave estimator must be evaluated in real room conditions to assess its benefit. Measurements require the use of a pressure-velocity (p-p) intensity measurement probe. Development of a cost-effective pressure-velocity (p-p) intensity measurement probe suitable for low-frequency applications is critical for real-world applications.

6 Conclusion

This paper proposes the theoretical framework for detecting and quantifying standing waves using a single point instantaneous energy measurement. The proposed standing wave estimator Γ leverages diffuseness metrics of energy density and temporal variation. Estimator Γ can detect the presence of standing waves, their directions, and magnitudes at all other locations except at the antinode of a standing wave. The diffuse field level relative to standing waves effects the estimator output. Future work will focus on validation in real rooms to assess the estimator's usefulness and the influence of diffuse field level at low frequencies.

References

[1] Braasch, J., Martens, W. L., and Woszczyk, W., "Modeling auditory localization of subwoofer signals in multi-channel loudspeaker arrays," in *Audio Engineering Society Convention 117*, Audio Engineering Society, 2004.

[2] Subkey, A., Cabrera, D., and Ferguson, S., "Localization and image size effects for low frequency sound," in *118th Convention of the AES, paper*, volume 6325, 2005.

[3] Mohamed, M. and Cabrera, D., "Human sensitivity to interaural phase difference for very low frequency sound," in *Acoustics 2008-Australian Acoustical Society Conference*, p. 47, 2008.

[4] Hill, A., Lewis, S., and Hawksford, M., "Towards a generalized theory of low-frequency sound source localization," in *Proceedings of the Institute of Acoustics*, volume 34, pp. 138–149, 2012.

[5] Hill, A. J. and Hawksford, M. O., "Subjective Evaluation of an Emerging Theory of Low-Frequency Sound Source Localization in Closed Acoustic Spaces," *Proc. IOA–Reproduced Sound 2014*, 36(pt 3), 2014.

[6] Martens, W. L., "Subjective evaluation of auditory spatial imagery associated with decorrelated subwoofer signals," in *Proc. 2002. International Conference of Auditory Display.*, 2002.

[7] Borenius, J., "Perceptibility of direction and time delay errors in subwoofer reproduction," in *Audio Engineering Society Convention 79*, Audio Engineering Society, 1985.

[8] Theile, G. and Kügler, C., "Loudspeaker reproduction: Study on the subwoofer concept," in *Audio Engineering Society Convention 92*, Audio Engineering Society, 1992.

[9] Welti, T. S., "Subjective comparison of single channel versus two channel subwoofer reproduction," in *Audio Engineering Society Convention 117*, Audio Engineering Society, 2004.

[10] Zacharov, N., Bech, S., and Meares, D. J., "The Use of Subwoofers in the Context of Surround Sound Program Reproduction," in *Audio Engineering Society Convention 102*, Audio Engineering Society, 1997.

[11] Kelloniemi, A., Ahonen, J., Paajanen, O., and Pulkki, V., "Detection of subwoofer depending on crossover frequency and spatial angle between subwoofer and main speaker," in *118th Convention of the AES, paper*, volume 6431, Citeseer, 2005.

[12] Rämö, J., Bergen, S., Parker, J., Yli-Kätkä, V.-M., and Pulkki, V., "Detection of Two Subwoofers: Effect of Broad-Band-Channel Level and Crossover Frequency," in *Audio Engineering Society Convention 132*, Audio Engineering Society, 2012.

[13] Nastasa, M., Pulkki, V., and Mäkivirta, A., "Auditory localisation of low-frequency sound sources," in *Audio Engineering Society Convention 154*, Audio Engineering Society, 2023.

[14] Nastasa, M., Pulkki, V., and Mäkivirta, A., "The Impact of Standing Waves on Localization Ability of Low-Frequency Sound Sources," *AES: Journal of the Audio Engineering Society*, 72(10), pp. 716–725, 2024.

[15] Jacobsen, F., "Sound intensity," *Springer Handbook of Acoustics*, pp. 1093–1114, 2014.

[16] Fahy, F., *Sound intensity*, CRC Press, 2017.

[17] Jacobsen, F., *The Diffuse Sound Field*, Ph.d. dissertation, Technical University of Denmark, 1979.

[18] Merimaa, J., *Analysis, synthesis, and perception of spatial sound: binaural localization modeling and multichannel loudspeaker reproduction*, Helsinki University of Technology, 2006.

[19] Nélisse, H. and Nicolas, J., "Characterization of a diffuse field in a reverberant room," *The Journal of the Acoustical Society of America*, 101(6), pp. 3517–3524, 1997.

[20] Ahonen, J. and Pulkki, V., "Diffuseness estimation using temporal variation of intensity vectors," in *2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 285–288, IEEE, 2009.

[21] Del Galdo, G., Taseska, M., Thiergart, O., Ahonen, J., and Pulkki, V., "The diffuse sound field in energetic analysis," *The Journal of the Acoustical Society of America*, 131(3), pp. 2141–2151, 2012.