DOI: 10.61782/fa.2025.0310

FORUM ACUSTICUM EURONOISE 2025

DOES SOURCE DIRECTIVITY INFLUENCE HUMAN PERCEPTION OF SPEECH POSITION?

Carla E. Contreras-Saavedra[1*] **Rhoddy Viveros-Muñoz**[2]
[1]Facultad de Ciencias de la Rehabilitación y Calidad de Vida, Universidad San Sebastián, Concepción, Chile.
[2]Department of Electronics and Computer Sciences, Universidad Técnica Federico Santa María, Concepción, Chile.

ABSTRACT

Humans rely on detecting and recognizing surrounding sounds, a complex task due to the variability, mobility, and contextual diversity of auditory stimuli. This study investigates human perception of speech source position—encompassing both direction of arrival (DOA) and distance—by addressing the question: "Does the directivity of a speech source influence human spatial perception?"
A listening test was conducted comparing omnidirectional and directional speech sources. Stimuli were generated using an acoustic virtual reality framework (RAVEN) under various conditions: two reverberation times (0.1 and 0.6 seconds), 8 DOA angles (0°, 30°, 90°, 150°, 180°, 210°, 270°, 330°), and two source types (omnidirectional and directional). Participants estimated the perceived direction and distance of the stimuli in a controlled environment.
Findings from this study enhance the understanding of how source directivity impacts spatial perception in humans, providing a benchmark for the performance of artificial neural networks in similar tasks. These insights have potential applications in the design of immersive auditory experiences, hearing aids, and spatial audio systems that bridge human perception and machine learning models. It also contributes to the understanding of the mechanisms that central auditory processing exerts to manage sound localization, which could have future clinical applications.

[*]***Corresponding author***: carla.contreras@uss.cl.

Keywords: *speech perception, direction of arrival, source directivity, acoustic virtual reality, source position estimation.*

1. INTRODUCTION

Human auditory perception enables us to detect, localize, and interpret sounds in complex environments, a capacity that is essential for communication, situational awareness, and spatial orientation. Accurately estimating the position of sound sources—commonly referred to as sound localization—relies on multiple acoustic cues, including interaural time and level differences (ITD and ILD), spectral filtering from the pinna, and environmental factors such as reverberation and background noise [1-4].
Among the wide number of auditory stimuli encountered in daily life, speech holds particular importance. Humans are highly attuned to vocal signals and depend on localizing speakers to engage in effective verbal interaction. Understanding how the human auditory system estimates both the direction of arrival (DoA) and the distance of speech sources is thus crucial for applications in auditory neuroscience, hearing technologies, and immersive audio design.
An often overlooked but critical characteristic of speech is its acoustic directivity—the way vocal energy radiates unequally across directions depending on articulation, frequency, and head orientation [5-6]. Most studies on sound localization assume idealized or omnidirectional source models, neglecting the impact that directional radiation patterns may have on the perception of spatial attributes, particularly in reverberant environments. This gap is especially relevant given that directivity can alter the

spectral and temporal cues available to the listener, potentially biasing distance or angular judgments.
To address this issue, the present study investigates whether the directivity of speech sources influences human spatial perception, specifically the estimation of DoA and distance. Using an acoustic virtual reality framework (RAVEN) [7], we simulated speech sources under two directivity conditions—omnidirectional and directional—across multiple DoAs and two levels of reverberation. Participants performed a listening task in which they estimated the perceived angular direction and distance of each stimulus.
We hypothesize that directional sources, due to their non-uniform energy radiation, will lead to systematic shifts in perceived direction and/or distance compared to omnidirectional sources, particularly in reverberant conditions. Additionally, we expect greater variability or error in spatial judgments when the directional source is not directly facing the listener.
By examining how directivity modulates spatial hearing in humans, this work contributes to a deeper understanding of central auditory processing mechanisms and offers valuable insights for improving models of artificial spatial perception. The results not only provide a perceptual benchmark for evaluating neural networks trained for spatial audio tasks but also hold implications for the development of hearing aids, clinical diagnostics, and immersive audio systems that aim to approximate or enhance human spatial hearing.

2. METHOD

2.1 Participants

Twenty normal-hearing adults (age range: 19–47 years; mean age: 29; 10 female) participated in the study. All participants reported no history of hearing disorders. Informed consent was obtained before participation, and all procedures followed the ethical guidelines of *Universidad de Santiago de Chile* ethics committee, protocol number 379/2023.

2.2 Stimuli and Experimental Conditions

The speech stimuli consisted of 700 phonetically balanced sentences derived from a Chilean Spanish adaptation of the SHARVARD corpus [8]. All sentences were recorded at a sampling rate of 44.1 kHz. A subset of these recordings was selected for spatialization.
Spatial rendering was performed using the RAVEN acoustic simulation framework, which allowed for precise control over room acoustics, source directivity and localization, and listener position. Three source directivity conditions were modeled:

- Omnidirectional, radiating uniformly in all directions.
- Directional-facing, using a human voice radiation pattern from RAVEN's directivity database, oriented toward the listener.
- Directional-reversed, using the same radiation pattern, but oriented away from the listener.

Stimuli were presented at eight DoAs: 0°, 30°, 90°, 150°, 180°, 210°, 270°, and 330°, and under two reverberation times: 0.1 s (low reverberation) and 0.6 s (moderate reverberation). Additionally, sources were placed at two fixed distances: 1.0 m and 3.0 m, allowing for an assessment of spatial perception across proximal and distal conditions.
Binaural renderings were created using measured HRTFs from the IHTA-indHARTF database [9], providing realistic spatial cues aligned with human anatomy and perception.

2.3 Apparatus and Listening Environment

The listening test was conducted in a sound-isolated booth. Participants wore high-fidelity headphones. Stimuli were presented using a custom PsychoPy script, and responses were collected through an on-screen spatial interface.
The virtual environment was visualized as a circle representing the horizontal plane around the listener, where participants could click to indicate the perceived DoA, and estimated distance of each sound source.

2.4 Procedure

Participants were seated in front of a computer screen and given a training session with corrective feedback prior to the experimental task. Each trial consisted of the playback of a single spatialized sentence. After hearing the sentence, participants used a graphical interface to indicate two perceptual estimates:

- the direction of arrival (DoA)
- the perceived distance of the source.

The test was designed to cover all combinations of 2 reverberation times (T30 = 0.1 s and 0.6 s), 8 azimuth angles (DoAs), 3 source directivities (facing the listener, facing away, omnidirectional), and 2 distances (1 m and 3 m), resulting in 96 unique experimental conditions. Each condition was repeated 8 times per participant to increase the number of observations, for a total of 768 trials per participant. The trial order was randomized to minimize order effects.

For the DoA response, participants used the numeric keypad of the computer, where each number corresponded to a specific azimuth: 8 = 0°, 9 = 30°, 6 = 90°, 3 = 150°, 2 = 180°, 1 = 210°, 4 = 270°, and 7 = 330°.
For the distance responses, participants used designated keys on the standard keyboard.
Participants were allowed to take short breaks after each trial to reduce fatigue. The total duration of the session was approximately 90 minutes per participant.
Design of statistical analysis: Two separate repeated-measures ANOVAs were conducted to examine the hypothesis. The first analysis evaluated the error rate (ER) in direction of arrival (DoA) estimation, expressed as the mean localization error as a function of the DoAs of the speech source (dependent variable). The within-subject factors included source directivity (facing the listener, facing away, omnidirectional), reverberation time (T30 = 0.1 s and 0.6 s), and azimuth angle (eight levels: 0°, 30°, 90°, 150°, 180°, 210°, 270°, and 330°).
The second analysis focused on error rate in distance estimation (mean localization error as a function of the distance of the speech source), again using repeated-measures ANOVA. The within-subject factors for this analysis were source directivity (facing the listener, facing away, omnidirectional), reverberation time (T30 = 0.1 s and 0.6 s), and distance (1 m and 3 m).

3. RESULTS

3.1 DoA Effect

For the first analysis, the ***mean localization error*** was calculated as a function of the DoAs of the speech source. The first trial of each experimental block was excluded from the data.
The measure was submitted to a repeated-measures ANOVA including source directivity, reverberation, and DoA.
The analysis revealed that source directivity did not have a significant main effect ($F(2, 19) = 1.22$, $p = .305$). Likewise, reverberation time did not reach statistical significance ($F(1, 19) = 2.26$, $p = .149$).
However, the analysis revealed a significant main effect of DoA ($F(7, 19 = 17.68$, $p < .001$). Mean localization errors followed a consistent pattern: the lowest errors were observed at 90° and 270°, while the highest errors occurred at 0°, 30°, 150°, and 180° (See Figure 1). These results suggest that participants were more accurate when localizing sources located on the lateral planes, likely due to the strong interaural cues present at those angles. In contrast, frontal and rear positions—particularly 0° and 180°—showed significantly greater error, consistent with known difficulties in resolving spatial cues along the sagittal plane.

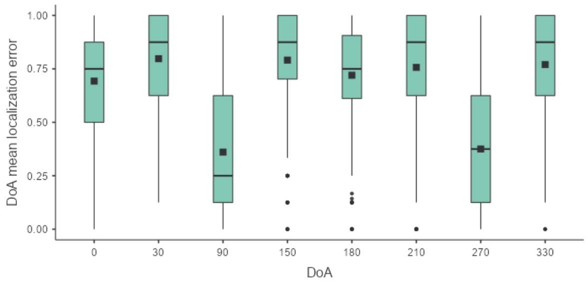

Figure 1. Mean Localization Error as a function of the DoAs of the speech source, based on participants' real spatial estimates.

A significant two-way interaction between DoA and source directivity was also observed ($F(17, 19) = 14.85$, $p < .001$). Across all directivity conditions, error rates were lowest at DoA 90° and 270°, indicating that lateral localization accuracy is robust across source configurations (see Figure 2). This reinforces the idea that spatial ambiguity increases for sources located along the sagittal axis.

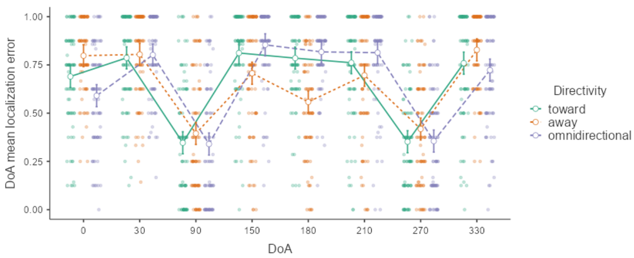

Figure 2. Mean Localization Error as a function of the two-way interaction of DoAs of the speech source with source directivity, based on participants' real spatial estimates.

The interaction between DoA and reverberation time was also significant ($F(7, 19 = 3.52$, $p = .002$).
Again, lower error rates were found at lateral positions (DoA = 90° and 270°) for both reverberation conditions. Under low reverberation (T30 = 0.1 s), error rates were particularly low at 90° (DoA 90° = 35.3%) and 270° (DoA 270° = 35.0%), compared to higher error rates at more frontal or rear positions (e.g., DoA 30° = 80.6%, DoA 150° = 77.6%). Under higher reverberation (T30 = 0.6 s), a similar pattern was observed, though with slightly increased error rates for some frontal directions (e.g., DoA 150° =

80.6%, DoA 210° = 78.3%). Lateral positions remained relatively less affected (DoA 90° = 36.8%, DoA 270° = 40.0%), suggesting that the impact of reverberation is greater for sources located near the frontal and rear axes, where binaural cues are typically less distinct.
These findings indicate that reverberation amplifies the difficulty of localizing sources at frontal and rear angles, while its effect is comparatively attenuated for lateral positions.

3.2 Distance Effect

For the second analysis, mean localization error was calculated as a function of the distance of the speech source. As in the previous analysis, the first trial of each block was excluded.
A repeated-measure ANOVA revealed a significant main effect of source directivity ($F(2, 19) = 11.63$, $p < .001$), with highest error rates found when the source was facing away from the listener (see Figure 3).
Reverberation time also had a significant main effect ($F(1, 19) = 12.94$, $p = .002$), with greater localization error observed under high-reverberation conditions (0.6 s).
Moreover, the two-way interaction between source directivity and reverberation time was significant, confirming that reverberation at 0.6 s consistently impaired distance estimation across all directivity conditions. The lowest error rates overall were found in the "toward" directivity condition, especially under low reverberation.

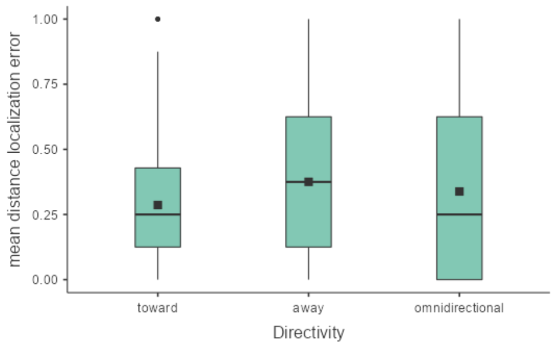

Figure 3. Mean Localization Error as a function of the distance of the speech source, based on participants' real spatial estimates.

Although distance did not show a main significant effect ($F(1, 19) = 2.70$, $p = .117$), the two-way interaction between source directivity and distance was statistically significant, indicating that localization error varied depending on the combination of these two factors (see Figure 4).

When the source was facing toward the listener, participants showed relatively low error rates at both distances, though error increased with distance: 0.194 m at 1 m, and 0.379 m at 3 m. In contrast, when the source was facing away from the listener, the pattern reversed: error was higher at 1 m (0.584 m) and notably lower at 3 m (0.167 m), suggesting an unexpected difficulty in estimating near-field sources in this condition. For omnidirectional sources, the difference across distances was especially marked: participants showed the lowest localization error at 1 m (0.089 m), but error increased substantially at 3 m (0.587 m).

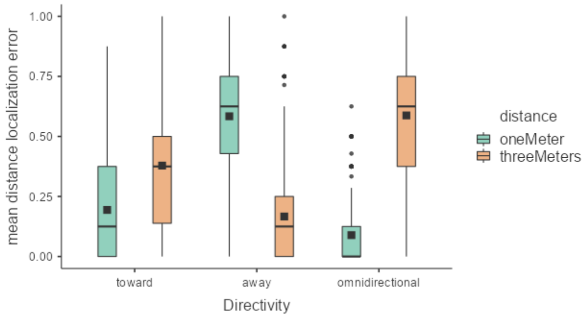

Figure 4. Mean Localization Error as a function of the two-way interaction of distance with directivity, based on participants' real spatial estimates.

This interaction suggests that the effect of distance on localization accuracy is not uniform across directivity conditions. Notably, errors increased with distance when the source faced the listener or was omnidirectional, but decreased with distance when the source faced away, possibly due to acoustic shadowing or altered binaural cues at closer distances in the "away" condition.
Also, the two-way interaction between distance and reverberation time was statistically significant, under low reverberation (T30 = 0.1 s), error rates were relatively low and similar across distances: 0.296 m at 1 m and 0.331 m at 3 m, showing only a slight increase with distance.
In contrast, under high reverberation (T30 = 0.6 s), the error increased more sharply with distance: from 0.282 m at 1 m to 0.424 m at 3 m. This suggests that greater reverberation amplifies the difficulty of accurately estimating farther distances, likely due to the masking or smearing of spatial acoustic cues in reverberant environments.

4. DISCUSSION

This study investigated how the directivity of speech sources affects human perception of spatial position,

defined by both direction of arrival (DoA) and distance. Our findings demonstrate that directivity, DoA, and their interaction with reverberation and distance all significantly influenced participants' ability to localize speech sources in reverberant environments.
Notably, reverberation time itself did not yield a significant main effect. This may reflect the limited range of reverberation times tested (0.1 s and 0.6 s), both of which fall within perceptually manageable boundaries for speech perception. It is possible that more extreme reverberant conditions could have yielded measurable effects, especially for more distant or rear-facing sources.
One of the key findings was that the interaction between source directivity and reverberation time had a significant effect on distance estimation. Highest errors were observed for sources facing away from the listener in highly reverberant conditions (T30 = 0.6 s), while forward-facing sources under low reverberation consistently yielded the lowest errors. This supports the hypothesis that the availability of direct-path acoustic cues, which are strongly attenuated when the source points away, plays a critical role in spatial perception. Interestingly, even omnidirectional sources yielded higher errors than forward-facing ones, suggesting that human listeners benefit from directional radiation patterns that emphasize the frontal acoustic field.
In terms of azimuthal location, a robust main effect of DoA was observed, with listeners were most accurate at identifying sources presented at 90° and 270°, consistent with optimal interaural time and level differences at lateral positions. In contrast, both frontal (0°, 30°) and rear (180°) positions resulted in significantly higher localization errors. This pattern is consistent with well-known front-back confusion effects, which arise from symmetrical binaural cues and reduced spectral contrast in the sagittal plane. Moreover, DoA interacted significantly with reverberation time, showing that errors at frontal and rear positions increased more under higher reverberation, whereas lateral positions remained relatively stable.
Unexpectedly, the interaction between source directivity and distance revealed an inverse effect for rear-facing sources: error decreased with distance in the "away" condition, suggesting possible perceptual recalibration or reduced cue conflict at longer ranges. In contrast, both forward-facing and omnidirectional sources showed increased error at greater distances, consistent with the expected degradation of spatial cues.
Although distance did not show a main effect, the interaction patterns with both directivity and reverberation confirmed its modulatory role. Specifically, reverberation had a greater detrimental effect on localization at 3 m than at 1 m, highlighting that spatial degradation becomes more pronounced as both source distance and environmental complexity increase.
Overall, our results suggest that directivity effects are not isolated but interact with spatial and acoustic variables such as distance and reverberation, modulating localization performance in complex ways. These findings have implications for the design of spatial audio systems, hearing aids, and virtual environments, where assumptions of omnidirectional speech radiation may underestimate the importance of realistic directivity cues. Furthermore, this dataset offers a valuable benchmark for comparing human localization performance with that of neural network-based models, advancing our understanding of both biological and artificial spatial hearing systems.

5. CONCLUSIONS

This study demonstrates that spatial localization accuracy is shaped by complex interactions between speech source directivity, azimuthal angle, reverberation time, and distance.
In particular, sources facing away from the listener consistently produced higher errors, especially in reverberant environments, emphasizing the importance of direct-path acoustic cues in spatial perception. Lateral source positions (90° and 270°) were associated with the highest localization accuracy across all conditions, while frontal and rear-central positions led to increased errors. Furthermore, reverberation disproportionately affects distance estimation for farther sources, amplifying localization difficulty under challenging acoustic conditions.
These findings offer valuable insight into the perceptual mechanisms underlying spatial hearing and highlight the need to incorporate realistic source directivity models in the design of hearing technologies, auditory simulations, and machine learning systems for spatial audio processing. Future work may further explore how these perceptual constraints compare with the performance of artificial neural networks and how source orientation cues can be leveraged to improve localization in real-world and assistive applications.

6. REFERENCES

[1] J. Blauert, *Spatial hearing: the psychophysics of human sound localization*. MIT press.

[2] E. B. Goldstein and J. R. Brockmole, *Sensation and perception*, Pacific Grove., vol. Vol. 90. CA, USA.: Wadsworth-Thomson Learning.

[3] A. W. Bronkhorst and R. Plomp, ‘The effect of head-induced interaural time and level differences on speech intelligibility in noise’, *J. Acoust. Soc. Am.*, vol. 83, no. 4, pp. 1508–1516, 1988.

[4] J. F. Culling, M. L. Hawley, and R. Y. Litovsky, ‘The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources’, *J. Acoust. Soc. Am.*, vol. 116, no. 2, pp. 1057–1065, Aug. 2004, doi: 10.1121/1.1772396.

[5] M. Noisternig, F. Zotter, and B. F. Katz, ‘Reconstructing sound source directivity in virtual acoustic environments’, *Princ. Appl. Spat. Hear.*, pp. 357–372, 2011.

[6] S. Pelzer, M. Pollow, and M. Vorländer, ‘Auralization of a virtual orchestra using directivities of measured symphonic instruments’, 2012.

[7] D. Schröder and M. Vorländer, ‘RAVEN: A real-time framework for the auralization of interactive virtual environments’, *Forum Acusticum*, pp. 1541–1546, 2011.

[8] V. Aubanel, M. L. G. Lecumberri, and M. Cooke, ‘The Sharvard Corpus: A phonemically-balanced Spanish sentence resource for audiology’, *Int. J. Audiol.*, vol. 53, no. 9, pp. 633–638, Sep. 2014, doi: 10.3109/14992027.2014.907507.

[9] F. Pausch, S. Doma, and J. Fels, ‘Hybrid multi-harmonic model for the prediction of interaural time differences in individual behind-the-ear hearing-aid-related transfer functions’, *Acta Acust.*, vol. 6, p. 34, 2022, doi: 10.1051/aacus/2022020.