JASA EXPRESS LETTERS ARTICLE asa.scitation.org/journal/jel

Speech intelligibility in multitalker situations with vibrotactile spatial cueing

Lilian Nguyen,[1,2] Gabriel Arnold,[2] Guillaume Andéol,[1] and Vincent Isnard[1,a)]
[1]Institut de Recherche Biomédicale des Armées, Brétigny-sur-Orge, 91220, France
[2]Caylar SAS, Villebon-sur-Yvette, 91140, France
lilian.nguyen91@gmail.com, gabriel.arnold@caylar.net, guillaume.andeol@def.gouv.fr, vincent.isnard@def.gouv.fr

Abstract: Degraded speech intelligibility in multitalker situations can be improved by spatial unmasking, using sound spatialization techniques such as binaural synthesis. However, intelligibility also depends on the ability to focus efficiently on the target. Three experimental sessions explored the benefit of an additional vibrotactile cue, spatialized around the waist, on intelligibility in different spatialized multitalker situations. Results indicate improvements in intelligibility scores and reduced listening effort specifically for an off-center target among masker talkers. Multimodality allows us to better understand the mechanisms of auditory attention and to open up new perspectives for improving speech intelligibility in multitalker situations.

[Editor: Peggy Mok] https://doi.org/10.1121/10.0036850

Received: 15 March 2025 **Accepted:** 23 May 2025 **Published Online:** 6 June 2025

1. Introduction

In multitalker situations, multiple speech signals interfere both perceptually and cognitively, affecting the intelligibility of the target message. Energetic masking results from acoustic interferences at the sensory level,[1,2] while informational masking arises from confusions between competing speech streams.[3] In this context, spatial separation of talkers was promptly identified as a major solution to improve intelligibility.[4–6] Indeed, "spatial unmasking" is supported by both the distribution of the acoustical signal on each ear and differentiated attention focused towards each talker direction.[7–9] In particular, Kidd *et al.*[9] measured higher speech intelligibility of a sentence masked by two other talkers when the target location was *a priori* known by participants. Similarly, Allen *et al.*[10] highlighted the existence of an auditory spatial gradient of attention, with an increase in unmasking as the target location was close to the attended location.

Furthermore, for colocated talkers, the attentional cost may lead to an increase in the associated listening effort.[11,12] By spatially separating a target talker from a masker in a speech intelligibility task, Andéol *et al.*[13] found reduced prefrontal activity, indicative of lower cognitive load and therefore listening effort, using functional near-infrared spectroscopy as a non-invasive brain imaging technique. Talker spatialization therefore appears beneficial for both intelligibility performance and listening effort.

Nonetheless, auditory localization abilities depend on temporal and spectral cues, produced by the filtering of the sound signal by the head and the ear pinna, depending on the direction of the sound source. In binaural synthesis, the head-related transfer functions (HRTFs) integrate these auditory spatialization cues allowing us to reproduce spatial auditory perception through headphones.[14] However, auditory localization abilities present limitations, both due to human perception and technical constraints, which may impact intelligibility in the case of spatialized speech signals.

The first factor of ambiguity in the localization of a sound source is the phenomenon of front-back confusion. The binaural cues of sounds at opposed positions relative to the interaural axis can be confused because of the symmetry relative to the ears, and generate an inversion in their localization between the front and the back.[15] As a result, the hindrance of spatial unmasking of sound sources, mistakenly perceived as colocated, can lead to a decrease in intelligibility. Moreover, virtual auditory displays are likely to generate more front-back reversals than in free-field conditions,[16] and in particular for non-individualized HRTFs.[17] Second, speech intelligibility is reduced for a target closed from a masker compared to higher angle spacing conditions.[8,18] Moreover, compared to free-field sources, the non-individualization of HRTF filters induces a localization blur which can impact intelligibility.[19] Third, target intelligibility may depend on its position relative to the maskers. Brungart and Simpson[20] reported generally better intelligibility when the target was at the leftmost or the rightmost location, due to the intensity ratio between the target and the maskers in each ear.

a)Author to whom correspondence should be addressed.

To overcome those limitations, a solution could be to reinforce the localization of the speech source of interest with an additional spatial cue, which would guide auditory attention and increase spatial unmasking. The tactile modality appears to be a good candidate to promote multimodal integration, which has the advantage of improving accuracy compared to unimodality.[21] Many studies have shown the efficiency of a tactile signal for orientation.[22–24] Moreover, tactile cueing of direction around the waist is intuitive as the signal is directly transmitted on the body and spatially congruent to the egocentric spatial map. Brill *et al.*[25] showed better auditory localization abilities in the horizontal plane with vibrotactile compared to auditory spatial cueing. Thus, a vibrotactile spatial cue, paired with the sound spatial cues, is expected to improve the localization of the target talker, and consequently its intelligibility over masking talkers.

As such, in the context of the limitations of auditory localization abilities in binaural synthesis, we explored the benefit of vibrotactile spatial cueing on speech intelligibility, as an additional contribution to direct auditory attention toward the target talker. Intelligibility scores were assessed for speech-in-speech tasks with spatialized talkers, without and with an additional vibrotactile spatial cueing of the target (i.e., audio-only vs audiotactile modality conditions). Three experimental sessions were designed in the azimuth plane and based on the paradigmatic cases presented previously: (1) front-back confusion, (2) angular spacing between talkers, and (3) the position of the target in the talkers group.

2. Methods

2.1 Participants

Twenty-eight French participants took part in this experiment and were compensated 20€ (13 women and 15 men; mean age 31.2 ± 8.6). None declared sensory pathology. The Speech, Spatial and Qualities of Hearing 15-item questionnaire (15iSSQ) was administered to check the hearing status.[26] For each part of the 15iSSQ (i.e., "Speech Hearing," "Spatial Hearing," and "Hearing Quality"), all participants' scores were above the mean minus two standard deviations. Moreover, all were normal-hearing as assessed by pure-tone audiometry, with hearing thresholds less than or equal to 25 dB HL at octave frequencies between 0.125 and 8 kHz (audiograms performed with an Echodia® Elios audiometer).

2.2 Materials

The speech materials consisted of 2048 sentences from the French version of the Coordinate Response Measure (CRM) corpus.[27] The sentences were structured as: "Prêt [call sign], va au point [color] [number] go" (French version of the original sentence: "Ready [call sign], go to [color] [number] now"). They were spoken by four male and four female voices, with 256 combinations of eight call signs (Alpha, Charlie, Delta, Echo, Kilo, Oscar, Tango, Whisky), four colors (Bleu, Vert, Rouge, Jaune), and eight numbers (from 1 to 8). For all experimental conditions and each trial, all sentence parameters were chosen randomly from eight talkers, eight call signs, four colors, and eight numbers, with the only conditions of having none of these parameters identical between the simultaneous sentences, and that one of them contains the call sign "Delta," corresponding to the target. The sentences had an average duration of 2.2 ± 0.1 s and were normalized by the root mean square (RMS) level.

To spatialize talkers in binaural synthesis, the sound files were convolved with a set of HRTF measurements taken from a KEMAR dummy head.[14] Only their azimuth was varied, while their elevation was fixed at 0° and their distance at 1.4 m. The mixture of spatialized sentences was normalized by the RMS level and broadcast through Beyerdynamic DT-770 Pro headphones (Beyerdynamic, Heilbronn, Germany), connected to a RME Fireface UCX-II sound card (RME Audio, Haimhausen, Germany), at a 44.1 kHz sampling rate. The sound level of the mixture was set to around 62 dB sound pressure level (SPL), as measured with a Brüel & Kjaer type 2250 sound level meter (Brüel & Kjaer, Naerum, Denmark) and a Brüel & Kjaer type 4153 artificial ear connected to a G.R.A.S. Type 12AK power supply.

A vibrotactile belt (Caylar®), equipped with 11 vibrotactile modules, was used for tactile stimulation around the participants' waist. Each module, covering an area of $20 \times 15\,mm^2$, was composed of four rotating mass vibrators with a cylindrical shape (5 mm in diameter, 11 mm long) and orthogonally disposed to the body. The vibrotactile belt was tied over the single-layer clothing for the entire duration of the experiment. Different sizes of stretch belts, on which the positions of the modules were adjustable, were available to fit participants' waist sizes. In order to distribute the modules in 30° steps from 0° (except 180° on the spine), the modules at 0° and ±90° were placed first, before placing the other modules equidistantly. The vibrotactile stimulation consisted of a series of 5 vibrations with a frequency of 150 Hz and a duration of 200 ms, separated by 200 ms, i.e., a total duration of 1800 ms. Its intensity was adjusted beforehand by the experimenters to a comfortable level while being easily detectable under laboratory conditions. Moreover, in the audiotactile condition, the vibrotactile stimulation was always emitted in the direction of the target talker, starting simultaneously and for approximately the same duration.

To complete the task, the participants were placed in a double-walled IAC audiometric booth. The experiment was programmed on MATLAB R2021a (Mathworks) and executed on a Lenovo ThinkPad L13 Yoga Gen2 laptop computer (Lenovo, Hong Kong). A graphical interface was displayed directly on the computer's touch screen. To respond, participants had to press one of the buttons of a 4×8 matrix of colored buttons, with four rows for the four colors and eight columns for the numbers indicated on each button from 1 to 8.

2.3 Spatial configurations of talkers

In the "front-back" session, the target talker was positioned either in front at −30° (left side) or +30° (right side), or behind at ±150° [Fig. 1(a)]. An ipsilateral masker talker was either colocated with the target or opposed relative to the interaural axis, resulting in eight spatial configurations of the talkers.

In the "maskers spacing" session, two masker talkers were presented on each side of the target, which was positioned either at 0°, −30°, or +30° [Fig. 1(b)]. The maskers were separated from the target by the same angle of either 30° or 60°, resulting in six spatial configurations.

Finally, in the "target position" session, three talkers were spaced at 30° intervals [Fig. 1(c)]. The target was one of these three talkers, positioned in the center or on one edge of the group, while the other two talkers were the maskers. As for the "maskers spacing" session, the group of talkers was positioned at 0°, −30°, or +30°. However, the ±60° target positions were not tested to shorten the session and maintain positions comparable to the other two, resulting in seven spatial configurations.

2.4 Procedure

The CRM task consisted of presenting, on each trial, a mixture of spatialized sentences. The instruction given to the participant was to focus on the sentence including the call sign "Delta," uttered by the target talker. At the end of the speech mixture, they had to report the color and number associated with this call sign by pressing the corresponding response button on the touch screen with no time limit. The other sentences were to be ignored. Participants were also instructed that, in the audiotactile condition, the vibrotactile stimulation indicated the target talker direction. If they did not know what answer to give, they had to press a button at random. The next trial started 0.5 s after their response.

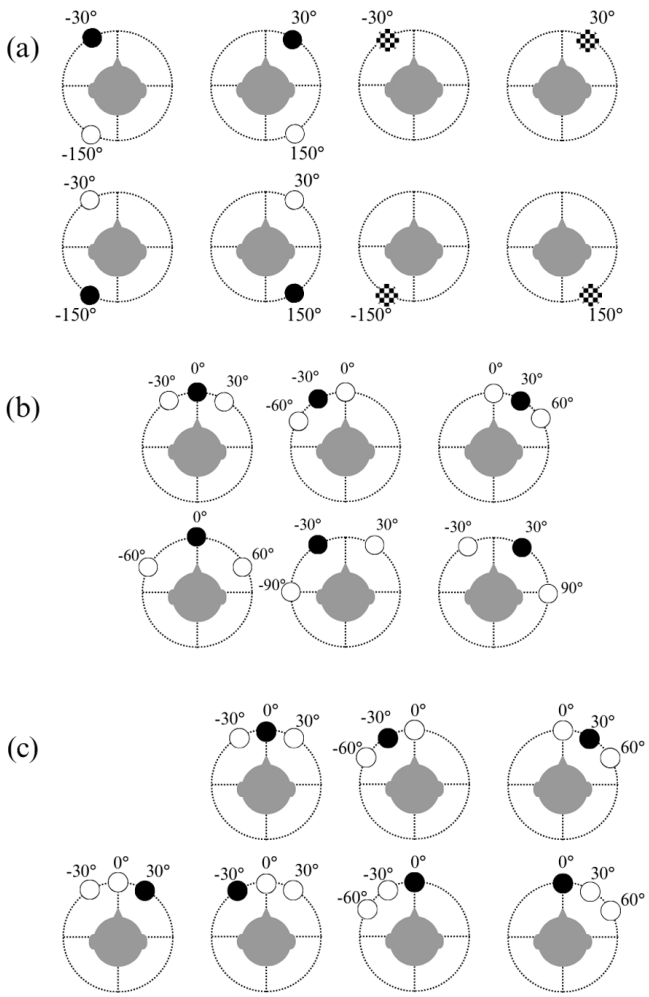

Fig. 1. Spatial configurations of the three experimental sessions: (a) "front-back"; (b) "maskers spacing"; (c) "target position." Black dots represent the target talker, white dots the maskers, and grid dots both when colocated.

At the beginning of the experiment, participants performed 24 trials without masker to familiarize themselves with the stimuli and the task: 12 trials in the audio-only condition, then 12 trials in the audiotactile condition. If the participant achieved 100% correct responses, the test phase could begin; otherwise, they repeated until achieving 100% correct responses. For a misunderstanding of the instructions, one participant repeated the familiarization phase.

The test phase included the three experimental sessions, whose order was counterbalanced across participants using a Latin square design. For each session, participants performed the task in both audio-only and audiotactile conditions, in an order that was also counterbalanced across participants. Note that, for each participant, this modality order was the same for all three sessions. Moreover, before each of the six experimental blocks, a short training consisted in presenting one random trial per spatial and modality conditions, i.e., respectively, 16, 12, and 14 trials for the "front-back", "maskers spacing" and "target position" sessions. In these training blocks only, feedback indicated the correct response. Then, the main blocks consisted of 16 trials per spatial and modality conditions, i.e., respectively, 256, 192, and 224 trials for the "front-back", "maskers spacing," and "target position" sessions, without feedback. A short break was programmed every 32 trials for the "front-back" and "maskers spacing" sessions, and every 28 trials for the "target position" session.

Finally, to evaluate listening effort, at the end of each of the six experimental blocks, participants responded on the screen to the Effort Scaling Categorical Unit (ESCU), by indicating the level of listening effort during the speech intelligibility task. This scale consisted of 14 graded categories ranging from "no effort required" to "only noise."[28]

2.5 Statistical analyses

For the three experimental sessions, data analyses were conducted on speech intelligibility scores and listening effort evaluations (from 1 to 14). For speech intelligibility, the percentage of correct responses was calculated for each experimental condition, considering only the correct responses for both color and number. However, for three participants, data from one session were excluded due to instruction misunderstanding for one participant, and malfunctions of the vibrotactile belt for two participants. Data of three other participants were excluded due to a high number of errors (for the three sessions for one, and one session for the other two). Indeed, their intelligibility scores were lower than the mean score minus two standard deviations of all the participants. Finally, one listening effort assessment was not collected due to a technical problem.

Repeated-measures analyses of variance (ANOVAs) were conducted for each experimental session. Whenever the sphericity assumption was violated, the degrees of freedom were reported using the Greenhouse–Geisser correction. Bonferroni *post hoc* tests were conducted when the main effects or interactions were significant. Paired Wilcoxon signed-rank tests were used for listening effort comparisons. A statistical test was considered significant when the p-value was less than 0.05. Statistical analyses were performed using JASP software version 0.19.3.

3. Results

3.1 Front-back session

An ANOVA was conducted on intelligibility scores with sensory modality (audio-only or audiotactile), talkers spacing (colocated or opposite), front-back hemifield of the target (front or back), and lateral hemifield (left or right), as within-participants factors. First, as might be expected, a significant main effect of talkers spacing indicated that the percentage of correct responses in the opposite condition was higher than in the colocated condition [$F(1, 25) = 37.15$, $p < 0.001$, $\eta_p^2 = 0.60$]. In addition, the main effects of front-back and lateral hemifields were not significant [respectively, $F(1, 25) = 3.45$, $p = 0.08$, $\eta_p^2 = 0.12$; $F(1, 25) = 3.12$, $p = 0.09$, $\eta_p^2 = 0.11$]. The analyses neither revealed a main effect of sensory modality [$F(1, 25) = 2.74$, $p = 0.11$, $\eta_p^2 = 0.10$] nor, contrary to what was expected, any interaction between sensory modality and talkers spacing [$F(1, 25) = 0.02$, $p = 0.88$, $\eta_p^2 = 0.00$]. Note that scores were already high, reaching 100% correct responses for several participants, suggesting a ceiling effect. Furthermore, although the three-way interaction between sensory modality, talkers spacing and lateral hemifield was significant [$F(1, 25) = 7.37$, $p < 0.05$, $\eta_p^2 = 0.23$], the differences between the audio-only and audiotactile conditions were not significant for each of the four talkers spacing and lateral hemifield conditions (Fig. 2). Finally, the mean listening effort level on this experimental session was 8.79 ± 2.11, without significant difference depending on the sensory modality [$p = 0.81$].

3.2 Maskers spacing session

An ANOVA was conducted on intelligibility scores with sensory modality (audio-only or audiotactile), maskers spacing (30° or 60°), and group position (0°, –30°, or +30°), as within-participants factors. As expected, the results showed a main effect of the maskers spacing, with a higher percentage of correct responses for a 60° compared to a 30° angle from the target [$F(1, 24) = 15.22$, $p < 0.001$, $\eta_p^2 = 0.39$]. There was also a main effect of the group position [$F(1.79, 42.99) = 13.48$, $p < 0.001$, $\eta_p^2 = 0.36$], with a significantly higher percentage of correct responses when the group was centered compared to ±30° [$p < 0.001$], and not between –30° and +30° [$p = 0.27$]. Furthermore, the analyses revealed a significant main effect of the sensory modality [$F(1, 24) = 9.24$, $p < 0.01$, $\eta_p^2 = 0.28$]. However, contrary to what was expected, the mean of correct responses in the audio-only condition was higher than in the audiotactile condition (Fig. 2). However, in this session, vibrotactile stimulation could not be used to disambiguate the target location, as the target was

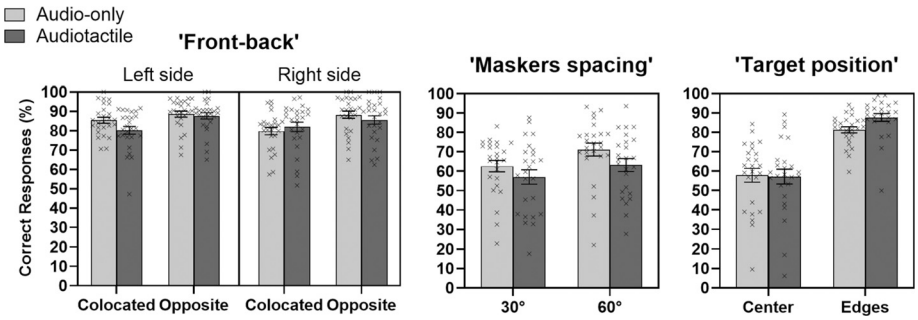

Fig. 2. Intelligibility scores for the three experimental sessions. Error bars represent the standard errors of the means and the points represent individual data.

systematically located at the center of the group (Fig. 1). Therefore, it may have acted as a disruptor instead of a useful cue, explaining the better score obtained in the audio-only condition. Finally, the analyses between the sensory modalities did not reveal a significant effect on the reported listening effort [$p = 0.69$], for a mean listening effort level of 10.30 ± 1.80.

3.3 Target position session

Similarly, an ANOVA was conducted on intelligibility scores with sensory modality (audio-only or audiotactile), target position (center or edges), and group position (–30°, 0°, or +30°), as within-participants factors. The results showed a main effect of target position [$F(1, 24) = 55.37$, $p < 0.001$, $\eta_p^2 = 0.70$]. When the target was on one of the edges of the talkers group, intelligibility scores were much higher compared to when it was in the center position, with a gain of around 27% points. The main effect of group position was not significant [$F(1.61, 38.59) = 2.79$, $p = 0.08$, $\eta_p^2 = 0.10$]. However, the analyses revealed a significant interaction effect between target and group positions [$F(1.54, 37.01) = 15.13$, $p < 0.001$, $\eta_p^2 = 0.39$], indicating that, for a target in the middle of the group, intelligibility scores were lower when the group was off-center vs centered [between left or right and center: $p < 0.01$; between left and right: $p = 1.00$]; while the opposite effect was observed for a target on the edges of the group, with scores higher when the group was off-center vs centered [between left and center: $p = 0.06$; between right and center: $p < 0.01$; between left and right: $p = 1.00$]. Moreover, this experimental session led to a significant main effect of sensory modality [$F(1, 24) = 4.40$, $p < 0.05$, $\eta_p^2 = 0.16$], with higher intelligibility scores in audiotactile than in audio-only condition. Actually, the interaction between sensory modality and target position was significant [$F(1, 24) = 7.57$, $p < 0.05$, $\eta_p^2 = 0.24$], with a gain of around 6% points for the audiotactile over the audio-only condition at the edges but not at the center positions [respectively, $p < 0.001$; $p = 1.00$] (Fig. 2). Finally, the benefit of vibrotactile cueing also extended to the reported level of listening effort. Indeed, the results showed a main effect of modality on listening effort [$p < 0.05$], which was rated significantly lower in the audiotactile condition (9.63 ± 1.81) than in the audio-only condition (10.32 ± 1.65).

4. Discussion and conclusion

This study replicated the main known effects of spatial unmasking on speech intelligibility, with better intelligibility for (1) the opposite condition compared to the colocated condition,[29] (2) a wider angular separation between the target and the masker talkers,[29,30] (3) a target on the edges of a group of talkers rather than in the center.[20] However, no systematic advantage was observed of vibrotactile spatial cueing, both in terms of intelligibility scores and listening effort.

In the "front-back" session, intelligibility was already high, even in the more difficult colocated condition, leaving little room for improvement for the audiotactile condition when talkers were spatially separated.[27] Moreover, as the vibrotactile cue indicated the direction of both target and masker in half of the trials (i.e., the colocated condition), participants may have disregarded it without impacting the cognitive resources available to perform the task.[31]

In the "maskers spacing" session, intelligibility scores were unexpectedly lower in the audiotactile condition. Since the target was always in the center of the group of talkers, the auditory attentional beam had to spontaneously focus on it. It, therefore, appears that rather than providing a useful cue for the task, the vibrotactile stimulation had a counter-productive distracting effect without however generating a difference in listening effort.[10]

In the "target position" session, intelligibility was first strongly improved when the target was located at the edges of the talkers group. Nevertheless, an additional benefit of vibrotactile spatial cueing was revealed specifically in this condition. It is suggested that two adjacent masker speech streams could be merged into a single stream, thus generating less interference with target processing.[4] It may also be explained by object-centered attention, which implies that spatial cueing is more effective in directing attention to a spatial part of a defined object than an absolute direction.[32] Moreover, vibrotactile cueing also reduced the associated listening effort, and this gain is certainly underestimated in the edge

condition because it was only assessed at the end of the modality condition blocks, and not between all experimental conditions.

A higher benefit of vibrotactile cueing should have been obtained if it had been presented before the auditory stimulus, indicating where to attend the target.[9] However, given the study's application field (e.g., radio communications), this would have been unrealistic. Limitations of the benefit of vibrotactile cueing may also arise from spatial misalignments between auditory and tactile stimuli, impacting multisensory integration.[33] Indeed, an auditory localization bias, depending in particular on the position of the sound source in space, can reach several degrees of azimuth.[34] While Van Erp[22] reported shifts up to 10° in judgment of azimuth directions cued with vibrotactile signals on the torso. Also, the authors observed a localization bias that can increase with increasing target azimuth angle and head orientation, in both auditory and tactile modalities, suggesting a disruption of internal spatial representation due to changes from the midline body references.[35]

Finally, the disparity in individual data when the task was made more difficult suggests varying auditory focusing abilities across participants. Target talker localization performance, and thus supposedly intelligibility, could be improved by using individualized rather than generic HRTFs, as well as by providing auditory localization training.[36] Similarly, enhancing temporal correspondence between modalities could support intelligibility in multitalker situations, for example using speech-derived vibrotactile signals.[37]

In conclusion, the present study shows that using vibrotactile spatial cueing can improve speech intelligibility in specific multitalker situations but may also be counterproductive in others, depending at least on the spatial configuration of the talkers.

Acknowledgments

This research was funded by Grant No. Biomedef HUM-1–2317. We thank Jean-Christophe Bouy and Paul-Antoine Boyer for their scientific and technical assistance.

Author declarations

Conflict of interest

Author Gabriel Arnold was employed by the company Caylar. The remaining authors state that they have no conflicts of interest to disclose.

Ethics approval

The Local Research Ethics Committee (IRB00013918) approved this specific study prior to the experiment under Ref. No. 2024C2E24, and all participants provided written informed consent to participate prior to any data collection.

Data availability

The data that support the findings of this study are openly available at https://doi.org/10.5281/zenodo.14939144.

References

[1]D. S. Brungart, B. D. Simpson, M. A. Ericson, and K. R. Scott, "Informational and energetic masking effects in the perception of multiple simultaneous talkers," J. Acoust. Soc. Am. **110**, 2527–2538 (2001).
[2]A. W. Bronkhorst, "The cocktail-party problem revisited: Early processing and selection of multi-talker speech," Atten. Percept. Psychophys. **77**, 1465–1487 (2015).
[3]G. Kidd, Jr., C. R. Mason, V. M. Richards, F. J. Gallun, and N. I. Durlach, "Informational masking," in *Auditory Perception of Sound Sources* (Springer, New York, 2008), Vol. 29, pp. 143–189.
[4]B. Arons, "A review of the cocktail party effect," J. Am. Voice I/O Soc. **12**, 35–50 (1992).
[5]R. Y. Litovsky, "Spatial release from masking," Acoust. Today **8**, 18–25 (2012).
[6]W. A. Yost, "Spatial release from masking based on binaural processing for up to six maskers," J. Acoust. Soc. Am. **141**, 2093–2106 (2017).
[7]W. A. Teder-Sälejärvi and S. A. Hillyard, "The gradient of spatial auditory attention in free field: An event-related potential study," Percept. Psychophys. **60**, 1228–1242 (1998).
[8]T. L. Arbogast, C. R. Mason, and G. Kidd, Jr., "The effect of spatial separation on informational and energetic masking of speech," J. Acoust. Soc. Am. **112**, 2086–2098 (2002).
[9]G. Kidd, T. L. Arbogast, C. R. Mason, and F. J. Gallun, "The advantage of knowing where to listen," J. Acoust. Soc. Am. **118**, 3804–3815 (2005).
[10]K. Allen, D. Alais, and S. Carlile, "Speech intelligibility reduces over distance from an attended location: Evidence for an auditory spatial gradient of attention," Percept. Psychophys. **71**, 164–173 (2009).
[11]J. E. Peelle, "Listening effort: How the cognitive consequences of acoustic challenge are reflected in brain and behavior," Ear Hear. **39**, 204–214 (2018).
[12]J. Rennies, V. Best, E. Roverud, and G. Kidd, Jr., "Energetic and informational components of speech-on-speech masking in binaural speech intelligibility and perceived listening effort," Trends Hear. **23**, 1–21 (2019).
[13]G. Andéol, C. Suied, S. Scannella, and F. Dehais, "The spatial release of cognitive load in cocktail party is determined by the relative levels of the talkers," J. Assoc. Res. Otolaryngol. **18**, 457–464 (2017).

[14]W. G. Gardner and K. D. Martin, "HRTF measurements of a KEMAR," J. Acoust. Soc. Am. **97**, 3907–3908 (1995).
[15]J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, UK, 1997).
[16]F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. II: Psychophysical validation," J. Acoust. Soc. Am. **85**, 868–878 (1989).
[17]G. D. Romigh and B. D. Simpson, "Do you hear where I hear?: Isolating the individualized sound localization cues," Front. Neurosci. **8**, 370 (2014).
[18]R. Drullman and A. W. Bronkhorst, "Multichannel speech intelligibility and talker recognition using monaural, binaural, and three-dimensional auditory presentation," J. Acoust. Soc. Am. **107**, 2224–2235 (2000).
[19]K. Kondo, T. Chiba, Y. Kitashima, and N. Yano, "Intelligibility comparison of Japanese speech with competing noise spatialized in real and virtual acoustic environments," Acoust. Sci. Technol. **31**, 231–238 (2010).
[20]D. S. Brungart and B. D. Simpson, "Improving multitalker speech communication with advanced audio displays" (Air Force Research Laboratory, Wright-Patterson AFB, OH, 2005).
[21]C. D. Wickens, J. Goh, J. Helleberg, W. J. Horrey, and D. A. Talleur, "Attentional models of multitask pilot performance using advanced display technology," in *Human Error in Aviation* (Routledge, London, UK, 2017), pp. 155–175.
[22]J. B. Van Erp, "Presenting directions with a vibrotactile torso display," Ergonomics **48**, 302–313 (2005).
[23]A. Cosgun, E. A. Sisbot, and H. I. Christensen, "Guidance for human navigation using a vibro-tactile belt interface and robot-like motion planning," in *Proceedings of the 2014 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 6350–6355 (2014).
[24]S. Schätzle and B. Weber, "Towards vibrotactile direction and distance information for virtual reality and workstations for blind people," in *Proceedings of the 9th International Conference of the Universal Access in Human-Computer Interaction (UAHCI)*, Los Angeles, CA (August 2–7, 2015), pp. 148–160.
[25]J. C. Brill, B. D. Lawson, and A. H. Rupert, "Audiotactile aids for improving pilot situation awareness," in *Proceedings of the 18th International Symposium on Aviation Psychology* (2015).
[26]A. Moulin, A. Pauzie, and C. Richard, "Validation of a French translation of the speech, spatial, and qualities of hearing scale (SSQ) and comparison with other language versions," Int. J. Audiol. **54**, 889–898 (2015).
[27]V. Isnard, V. Chastres, and G. Andéol, "French version of the coordinate response measure corpus and its validation on a speech-on-speech task," JASA Express Lett. **4**, 075203 (2024).
[28]M. Krueger, M. Schulte, T. Brand, and I. Holube, "Development of an adaptive scaling method for subjective listening effort," J. Acoust. Soc. Am. **141**, 4680–4693 (2017).
[29]D. Yao, J. Zhao, L. Wang, Z. Shang, J. Gu, Y. Wang, M. Jia, and J. Li, "Effects of spatial configuration and fundamental frequency on speech intelligibility in multiple-talker conditions in the ipsilateral horizontal plane and median plane," J. Acoust. Soc. Am **155**, 2934–2947 (2024).
[30]A. Westermann and J. M. Buchholz, "The influence of informational masking in reverberant, multi-talker environments," J. Acoust. Soc. Am. **138**, 584–593 (2015).
[31]V. Santangelo, C. Ho, and C. Spence, "Capturing spatial attention with multisensory cues," Psychonomic Bull. Rev. **15**, 398–403 (2008).
[32]M. Turatto, V. Mazza, and C. Umilta, "Crossmodal object-based attention: Auditory objects affect visual processing," Cognition **96**, B55–B64 (2005).
[33]P. Bruns, C. Spence, and B. Röder, "Tactile recalibration of auditory spatial representations," Exp. Brain Res. **209**, 333–344 (2011).
[34]S. Carlile, P. Leong, and S. Hyams, "The nature and distribution of errors in sound localization by human listeners," Hear. Res. **114**, 179–196 (1997).
[35]C. Ho and C. Spence, "Head orientation biases tactile localization," Brain Res. **1144**, 136–141 (2007).
[36]G. Andéol, S. Savel, and A. Guillaume, "Perceptual factors contribute more than acoustical factors to sound localization abilities with virtual sources," Front. Neurosci. **8**, 451 (2015).
[37]Y. Oh, N. Kalpin, J. Hunter, and M. Schwalm, "The impact of temporally coherent visual and vibrotactile cues on speech recognition in noise," JASA Express Lett. **3**, 025203 (2023).