



The effect of spatial energy spread on sound image size and speech intelligibility

Ahrens, Axel; Marschall, Marton; Dau, Torsten

Published in:
Journal of the Acoustical Society of America

Link to article, DOI:
[10.1121/10.0000747](https://doi.org/10.1121/10.0000747)

Publication date:
2020

Document Version
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

Citation (APA):
Ahrens, A., Marschall, M., & Dau, T. (2020). The effect of spatial energy spread on sound image size and speech intelligibility. *Journal of the Acoustical Society of America*, 147(3), 1368-1378.
<https://doi.org/10.1121/10.0000747>

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

The effect of spatial energy spread on sound image size and speech intelligibility

Axel Ahrens, Marton Marschall, and Torsten Dau

Citation: *The Journal of the Acoustical Society of America* **147**, 1368 (2020); doi: 10.1121/10.0000747

View online: <https://doi.org/10.1121/10.0000747>

View Table of Contents: <https://asa.scitation.org/toc/jas/147/3>

Published by the [Acoustical Society of America](#)

ARTICLES YOU MAY BE INTERESTED IN

[Prediction of individual speech recognition performance in complex listening conditions](#)

The Journal of the Acoustical Society of America **147**, 1379 (2020); <https://doi.org/10.1121/10.0000759>

[Speech detection and localization in a reverberant multitalker environment by normal-hearing and hearing-impaired listeners](#)

The Journal of the Acoustical Society of America **147**, 1469 (2020); <https://doi.org/10.1121/10.0000844>

[Indoor sound level spectra of public entertainment premises for rating airborne sound insulation](#)

The Journal of the Acoustical Society of America **147**, EL215 (2020); <https://doi.org/10.1121/10.0000800>

[Overview of geometrical room acoustic modeling techniques](#)

The Journal of the Acoustical Society of America **138**, 708 (2015); <https://doi.org/10.1121/1.4926438>

[Can background noise increase the informational masking in a speech mixture?](#)

The Journal of the Acoustical Society of America **147**, EL144 (2020); <https://doi.org/10.1121/10.0000719>

[Effect of the number of amplitude-compression channels and compression speed on speech recognition by listeners with mild to moderate sensorineural hearing loss](#)

The Journal of the Acoustical Society of America **147**, 1344 (2020); <https://doi.org/10.1121/10.0000804>

SUBMIT TODAY!

JASA
THE JOURNAL OF THE
ACOUSTICAL SOCIETY OF AMERICA

**Special Issue: Theory and
Applications of Acoustofluidics**

The effect of spatial energy spread on sound image size and speech intelligibility

Axel Ahrens,^{a)} Marton Marschall,^{b)} and Torsten Dau^{c)}

Hearing Systems Section, Department of Health Technology, Technical University of Denmark, Building 352, Ørstedes Plads, 2800 Kongens Lyngby, Denmark

ABSTRACT:

This study explored the relationship between perceived sound image size and speech intelligibility for sound sources reproduced over loudspeakers. Sources with varying degrees of spatial energy spread were generated using ambisonics processing. Young normal-hearing listeners estimated sound image size as well as performed two spatial release from masking (SRM) tasks with two symmetrically arranged interfering talkers. Either the target-to-masker ratio or the separation angle was varied adaptively. Results showed that the sound image size did not change systematically with the energy spread. However, a larger energy spread did result in a decreased SRM. Furthermore, the listeners needed a greater angular separation angle between the target and the interfering sources for sources with a larger energy spread. Further analysis revealed that the method employed to vary the energy spread did not lead to systematic changes in the interaural cross correlations. Future experiments with competing talkers using ambisonics or similar methods may consider the resulting energy spread in relation to the minimum separation angle between sound sources in order to avoid degradations in speech intelligibility. © 2020 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>). <https://doi.org/10.1121/10.0000747>

(Received 4 May 2019; revised 30 January 2020; accepted 30 January 2020; published online 2 March 2020)

[Editor: Virginia Best]

Pages: 1368–1378

I. INTRODUCTION

The concept of the perceived size of acoustic sources, often referred to as the apparent source width or the sound image size, was first discussed in the context of concert hall acoustics (see [Griesinger, 1997](#), for a review) but has since been adopted in other areas within acoustics. Unlike in vision, where the perceived size of a visual object is directly related to the size of its retinal image ([Hering, 1861](#); [Holway and Boring, 1941](#)), the perception of the size of an auditory object seems less straightforward. The sound image size has been shown to be affected by early reflections in a given environment and is thus related to the amount of reverberation in the environment ([Blauert and Lindemann, 1986a](#)). An increased amount of reverberation results in a decrease of the correlation between the signals at the left and right ears of a listener, i.e., a reduced interaural cross correlation (IACC), which has been linked to larger perceived sources (e.g., [Blauert and Lindemann, 1986b](#)). In listeners with a hearing impairment, it was found that the range of perceived sound image sizes is generally reduced compared to that observed in normal-hearing listeners ([Whitmer et al., 2012, 2014](#)). Other studies demonstrated that dynamic range compression in hearing aids or simulated hearing aids, reflecting a level-dependent amplification scheme commonly used to compensate for loudness

recruitment, leads to enlarged sound image percepts ([Wiggins and Seeber, 2011, 2012](#); [Hassager et al., 2017](#)).

While there is evidence that the acoustic environment, the transmission through a device like a hearing aid, as well as effects of hearing impairment can affect human listeners' perception of the sound image size, only a few studies investigated how such altered spatial perception affects speech intelligibility.

A link between the sound image size and speech intelligibility might be expected based on the fact that spatial differences between target speech and interferers in the horizontal plane ([Duquesnoy, 1983](#)) or vertical plane ([Martin et al., 2012](#)), as well as in terms of distance ([Westermann and Buchholz, 2015](#)), are advantageous for speech intelligibility relative to conditions with colocated sources. [Cubick et al. \(2018\)](#) investigated the effect of hearing-aid amplification on spatial release from masking (SRM) and the sound image size of the target and the interferers for normal-hearing listeners. They found larger sound image sizes, as well as a reduced SRM, in the conditions with hearing aids compared to the conditions without hearing aids but did not show a definitive link between the measures. In general, the extent to which point-like sound sources might be easier to perceptually segregate than large sound sources and how the sound image size is related to speech intelligibility in conditions with one or more interferers has not been studied systematically.

The spatial extent of a sound source reproduced over loudspeakers can be described by its spatial energy spread. Thus, controlling the energy spread of reproduced sound

^{a)}Electronic mail: aahr@dtu.dk, ORCID: 0000-0001-8800-2497.

^{b)}ORCID: 0000-0003-2534-7062.

^{c)}ORCID: 0000-0001-8110-4343.

sources may allow for a systematic manipulation of the sound image size. In the present study, the spatial energy spread of sound sources was varied using ambisonics processing, a method based on spherical harmonic decomposition (Gerzon, 1973). The higher the ambisonics order, the larger the number of spherical harmonic components and, thus, the smaller the spatial energy spread of the reproduced sources (Gerzon, 1992; Daniel, 2001; Bertet *et al.*, 2007; Zotter and Frank, 2012). While explicit source-widening algorithms exist, such as the one proposed by Zotter *et al.* (2014), here the choice was made to consider the effects of the ambisonics reproduction order directly, which might have implications for speech tests presented in virtual sound environments (e.g., Oreinos and Buchholz, 2016; Ahrens *et al.*, 2019). Thus, in the current study, the energy spread refers to a physical quantity that describes the spatial extent of a sound source reproduced in the virtual environment, whereas the analogous perceptual attribute is termed sound image size. The latter expression is favored over the “apparent source width” in order not to restrict the definition to describing only the horizontal extent.

Three experiments were conducted to investigate the effects of the spatial energy spread on speech intelligibility in young normal-hearing listeners. Experiment 1 explored to what extent the energy spread affects the corresponding (perceived) sound image by measuring the location and size of sound images of speech sounds as a function of their spatial energy spread. Experiment 2 investigated if speech intelligibility is affected by the energy spread in conditions with colocated and spatially separated target-interferer configurations. In experiment 3, the separation angle between the target and the interferers required to achieve a fixed level of speech intelligibility was estimated for different degrees of energy spread of the target and interferers. To analyze the potential perceptual cues that may contribute to the sound image size, a variation of the IACC, considering only the early reflections and three octave bands, was analyzed (Okano *et al.*, 1998; Frank, 2013).

II. GENERAL METHODS

A. Listeners

Thirteen young (20–27 year olds) normal-hearing listeners participated in the study. Six of the listeners carried out the spatial perception experiment (experiment 1), and all participated in the speech intelligibility experiment with spatially distributed interfering talkers fixed in space (experiment 2). Ten listeners participated in the speech intelligibility experiment with an adaptive spatial configuration of the interfering talkers (experiment 3), and six of them also participated in experiments 1 and 2. Thus, 6 of the 13 listeners participated in all 3 experiments. All listeners were native Danish speakers and paid on an hourly basis. Audiograms were measured for all listeners at the octave band frequencies between 250 Hz and 8 kHz. All thresholds were below or equal to 20 dB hearing level (HL).

The participants provided informed consent, and all experiments were approved by the Science-Ethics Committee for the Capital Region of Denmark (reference H-16036391). The order of the experiments was randomized for each listener. Single sessions were limited to a duration of 2.5 h, and the listeners were encouraged to take breaks during the sessions.

B. Virtual sound environment

All experiments were conducted in an anechoic chamber. The anechoic chamber was equipped with 64 KEF LS50 loudspeakers (KEF Audio, Maidstone, UK), arranged in a spherical array. In the current study, only the 24-loudspeaker horizontal ring at ear height was used. The height of the chair was individually adjusted for each listener. The 24 loudspeakers were equidistantly spaced on a 2.4 m radius (separation of 15°). The loudspeakers were driven by a sonible d:24 amplifier (sonible GmbH, Graz, Austria). The audio signals were generated in MATLAB (The Mathworks Inc., Natick, MA) and fed to the amplifier via a digital audio network through Ethernet (DANTE) and two TESIRA biamp DSP units, including TESIRA SOC-4 digital-to-analog converters (biamp Systems Inc., Beaverton, OR). Level, time, and frequency response corrections were applied, based on impulse response measurements at the midpoint of the loudspeaker array.

C. Stimuli and spatialization of sounds

The speech stimuli that were used throughout this study were taken from the multi-talker version of the Dantale II, a Danish matrix sentence test (Wagener *et al.*, 2003; Behrens *et al.*, 2007). The stimuli were spatialized using ambisonics reproduction on the horizontal 24-loudspeaker array. A 24-transducer setup allows for a maximum ambisonics order, M , of 11 (Gerzon, 1973). In addition to the 11th-order reproduction, 1st-, 3rd-, and 5th-order ambisonics were investigated using all 24 loudspeakers on the horizontal ring. The particular orders were chosen to cover the full range reproducible with the loudspeaker setup while focusing on lower orders where the absolute changes in the energy spread are larger.

To examine possible spectral distortions introduced by ambisonics reproduction at off-center positions (Solvang, 2008), an optimal subset of $N = 2M + 2$ loudspeakers (Daniel, 2001) was investigated for first-, third-, and fifth-order ambisonics (i.e., for $M = 1, 3, 5$). However, since no significant differences in the results obtained with the full set and the subset of loudspeakers were found, only the results for the full set are presented here.

The loudspeaker signals were generated using a dual-band decoder with a crossover frequency at $M \times 700$ Hz (Favrot and Buchholz, 2010). Below the crossover frequency basic ambisonics decoding was used, and above the crossover frequency “max r_E ” decoding was used (Daniel, 2001). The loudspeaker signals were presented to the listeners anechoically (direct sound only) and included simulated reverberation from a small, living-room type area [International Electrotechnical Commission (IEC) listening room; IEC 268-

13, 1985] with a volume of 100 m^3 and a reverberation time of about 0.4 s. The room was modeled using the room acoustics simulation software Odeon (Odeon A/S, Lyngby, Denmark) and is available online (Ahrens, 2018). The loudspeaker signals were generated using the LoRA toolbox (Favrot and Buchholz, 2010). Since only loudspeakers in the horizontal plane were employed, the elevated reflections were mapped to the horizontal plane. The simulated sources were placed at a distance of 2.4 m and therefore coincided with the distance of the loudspeaker array.

Ambisonics decoding at different orders can lead to variations in the frequency response due to the different decoder crossover frequencies as well as the spectral colorations when more than $2M + 1$ loudspeakers are used (Solvang, 2008). To reduce the influence of spectral colorations on the experimental outcomes, equalization filters were designed to achieve equal frequency responses as measured at the center of the loudspeaker array. The filters were designed to match the direct sound (anechoic) frequency response of the 11th-order ambisonics reproduction. The reverberant impulse responses were equalized with the same filters as the anechoic impulse responses. Subsequently, the impulse responses for both anechoic and reverberant conditions were set to the unity gain of the direct sound. Thus, the reverberant condition was perceived as somewhat louder than the anechoic condition, while the source levels remained equal.

The spatial energy spread of the virtual sources that were reproduced using ambisonics can be described using the ambisonics energy vector, r_E (Gerzon, 1992; Daniel, 2001). The angular energy spread is defined as the inverse cosine of the length of the energy vector (Daniel, 2001; Zotter and Frank, 2012; Bertet et al., 2013). For an infinite ambisonics order, the energy vector is equal to one, i.e., the energy spread is zero. For lower orders, the length of the energy

vector is reduced from one, and the energy spread increases. Figure 1 shows the ambisonics panning function of the ambisonics orders considered in the current study. The arrow indicates the length of the energy vector, which can be related to the physical energy spread in degrees, which is indicated by the cross (Zotter and Frank, 2012). The length of the energy vector has been shown to correlate with the perceived sound image size of pink noise in normal-hearing listeners (Frank, 2013). The ambisonics panning function was calculated and plotted using the spherical array processing toolbox (Politis, 2016). The circles in Fig. 1 indicate the -3 dB beamwidth, i.e., the angle at which a signal at 0° is attenuated by 3 dB.

D. Statistical analyses and data

The results obtained in the three experiments were analyzed employing linear mixed-effects models using the statistics software *R* and the step function included in the lmerTest package (Kuznetsova et al., 2014). If *post hoc* analyses of within-factor comparisons were performed, the “emmeans” package was used to estimate marginal means from the mixed-effects linear models (Lenth, 2016). The *p*-values are reported, including Bonferroni significance corrections.

The data from the three experiments are available online in the supplemental material.¹

III. EXPERIMENT 1: MEASURES OF THE PERCEIVED SOUND IMAGE SIZE AND LOCATION

A. Methods

The listeners were asked to localize a single sound source and judge the size of the perceived sound image. This was done by indicating the location and size of the perceived sound image on the touchscreen of a mounted 9.7 in. Apple iPad Air 2 (Apple Inc., Cupertino, CA). Figure 2

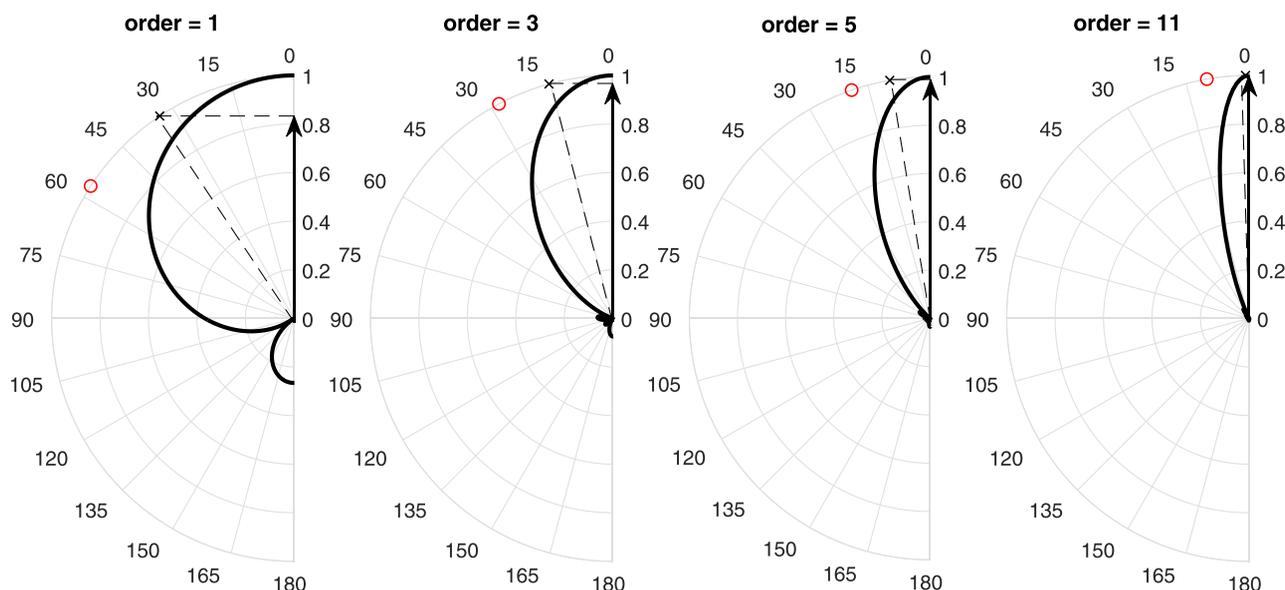


FIG. 1. (Color online) Ambisonics panning functions of different orders. The arrow indicates the length of the energy vector (r_E), and the cross represents the corresponding energy spread in degrees (calculated as the inverse cosine of the length of the ambisonics energy vector). The red circle indicates the 3 dB beamwidth of the panning function.

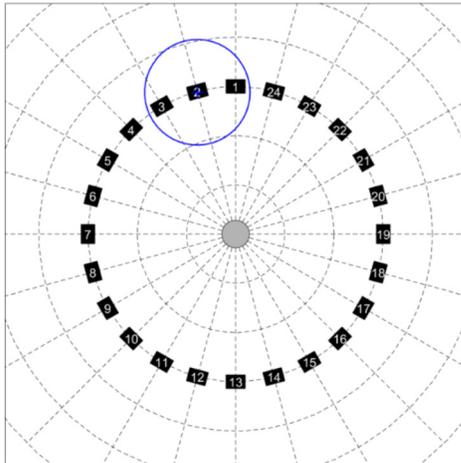


FIG. 2. (Color online) Screenshot of the user interface (UI) from the spatial perception experiment. The gray circle in the center depicts the listeners' position, and the black boxes indicate the loudspeaker locations. The numbers in the UI correspond to numbers displayed on the loudspeakers.

illustrates the user interface (UI) as shown to the listeners. To indicate the location of the sound image, the listeners were asked to place a cross at the desired location with a finger on the touchscreen. To indicate the size of the sound image, the listeners could vary the size of a circle around the cross by moving a finger closer to or further away from the origin, as in Hassager *et al.* (2017). The initial radius of the source size was randomized to reduce a potential bias. If multiple sound images ("split images") were perceived by the listeners, two or more circles could be placed on the UI. The listeners were instructed that sound images could be placed at any location and distance from the origin, i.e., also at positions closer to the listener than the loudspeaker ring or further away from it. The sound image size was defined as the area of the circle, and the source distance was defined as the length between the listener position and the center of the circle placed by the listener.

The sound sources were generated using different ambisonics orders and in conditions with and without simulated reverberation. The signal emitted by the sound source was either a single sentence spoken by a female talker from the Dantale II database or a speech-modulated noise (SMN) signal. The SMN had the same long-term spectrum and broadband envelope as the speech sentence but with random phase (Best *et al.*, 2013; Westermann and Buchholz, 2015; Ahrens *et al.*, 2019). The stimuli were either presented from the front (0°) or 15° azimuth to the right. Each condition, consisting of a given stimulus type (speech or SMN), location, reverberation, and ambisonics order was repeated 3 times, leading to 96 trials for each listener. The listeners were allowed to listen to each sound repeatedly before indicating the position and size of the sound image. Additionally, a reference sound was available to the listeners, providing an anchor with the minimum energy spread. The reference stimulus was generated using the same stimulus type as the target but was presented anechoically from a single loudspeaker in front of the listeners. The listeners could listen to the reference repeatedly and

were informed that the reference stimulus was of the smallest possible size.

The IACC was calculated from the first 80 ms of the binaural impulse responses (BIRs) and averaged over three octave bands at 0.5 kHz, 1 kHz, and 2 kHz. This measure is referred to as $IACC_{E3}$. The BIRs were measured in the center of the loudspeaker array using a B and K Head and Torso Simulator (type 4128-C; Brüel and Kjær A/S, Nærum, Denmark).

B. Results and discussion

Figure 3 shows the overlaid responses of all listeners obtained for the four ambisonics orders $M = 1$ (red, upper left panel), 3 (green, upper right panel), 5 (blue, lower left panel), and 11 (cyan, lower right panel) with both stimulus types presented from the front and lateral directions in the anechoic condition. Each semitransparent circle represents a single response. The size of the sound images did not seem to vary much across the conditions with different ambisonics orders. However, the position of the sound images was generally considered to be closer to the listener for low ambisonics orders than for the higher orders. In the following, the reported sizes and distances are considered in more detail. A separate analysis of the localization accuracy is not provided as only two source locations were employed in this study.

Figure 4 shows the perceived distance as a function of the ambisonics order in the anechoic (light gray) and reverberant (dark gray) conditions. The statistical analysis showed significance for all main effects [order, $F(3,561) = 9.2$, $p < 0.0001$; stimulus type, $F(1,561) = 4.1$, $p = 0.0442$; direction, $F(1,561) = 4.1$, $p = 0.0428$; reverberation, $F(1,561) = 210.9$, $p < 0.0001$] as well as the interaction between the ambisonics order and reverberation condition [$F(3,561) = 10.9$, $p < 0.0001$]. Thus, for the low orders, the anechoic sources were perceived to be closer to the listener than for the high orders. In fact, only the 11th-order condition was not perceived to be significantly further/closer to the actual loudspeaker distance at 2.4 m [$t(6.81) = -0.1$, $p = 0.09$]. The perceived distance for all other orders differed significantly from the actual distance ($p < 0.0167$) when presented anechoically. In the reverberant condition, none of the ambisonics orders led to a perceived distance that was significantly different from the actual loudspeaker distance ($p > 0.69$).

A comparison of the two spatial locations of the source showed that the lateral source was perceived, on average, to be 0.12 m closer to the actual loudspeaker location than the frontal source [$t(561) = -2.0$, $p = 0.0428$]. The comparison between the two stimulus types showed that the noise stimulus was perceived, on average, 0.11 m closer to the actual distance than the speech stimulus [$t(561) = 2.0$, $p = 0.0442$]. Thus, both differences are small with regard to the overall distance.

Figure 5 shows the perceived size (area in m^2) of the sound images as a function of the ambisonics order in the anechoic (top) and reverberant (bottom) conditions. A linear mixed model was fitted to the sound image size, where the ambisonics order, stimulus type, source location, and reverberation

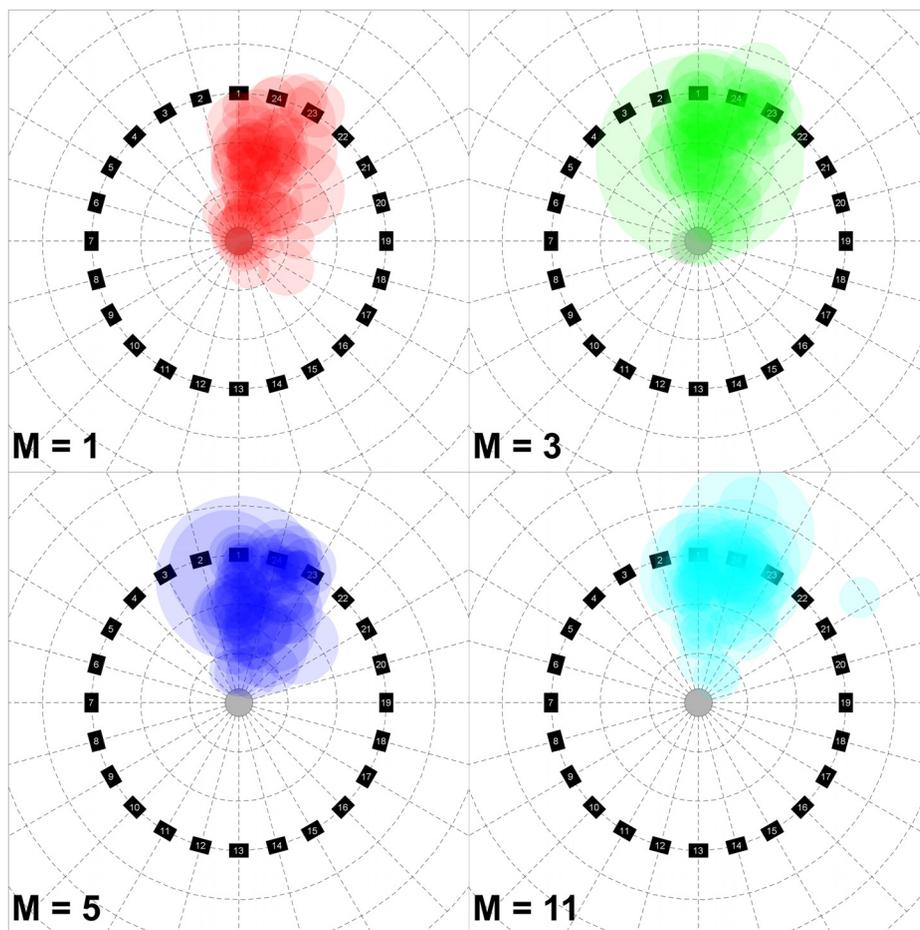


FIG. 3. (Color online) Overlay of the responses in the spatial perception experiment of all listeners with the speech and noise signals in the anechoic condition for sources from both 0° (front) and 15° (right). The signals were reproduced using the ambisonics orders (M) as shown in the subfigures.

condition were treated as fixed effects, and the effect of the listeners was treated as a random effect. The analysis of the model revealed that only the effect of reverberation contributed significantly to the model [$F(1,569)=102.4$,

$p < 0.0001$], while the other main effects, as well as the interactions, did not reach significance.

No differences between the sound image sizes obtained for the different ambisonics orders were found, even though larger sound images were expected for low orders as the spatial energy spread is larger with low orders, as shown in Fig. 1. In contrast to the current study, Frank (2013) found that the energy spread was highly correlated with the sound image size when a pink noise was presented over pairs or triplets of loudspeakers at various opening angles. Furthermore, Frank (2013) found a high correlation between the apparent source size and the $IACC_{E3}$. Figure 6 shows the $IACC_{E3}$ in the anechoic and reverberant conditions for sources from the front (0°) and side (15°) as a function of the ambisonics orders. While there are small variations, no clear trend can be seen with respect to the ambisonics orders. Thus, varying the energy spread by changing the ambisonics orders did not result in systematic changes in the $IACC_{E3}$, unlike with the method employed in Frank (2013). This may explain the lack of significant differences in the sound image size ratings in the current study.

Bertet et al. (2013) observed a lower localization precision (larger variance) with low ambisonics orders than with higher orders. The localization precision has been thought to be a measure of the sound image size percept (Blauert, 1984), however, Whitmer et al. (2014) did not find a correlation between localization precision and apparent source

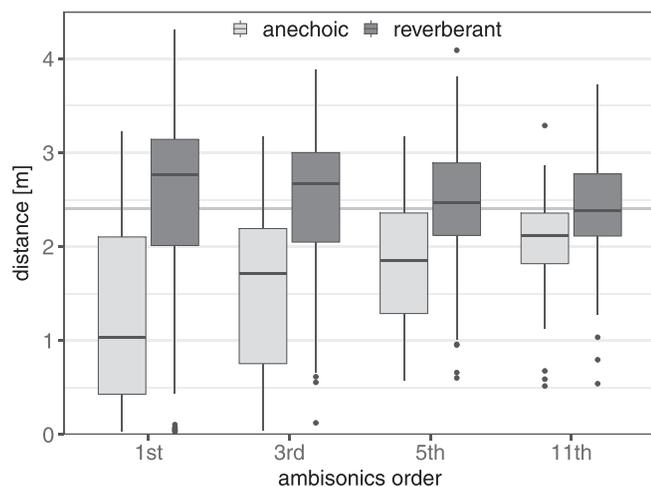


FIG. 4. Perceived distances of the speech and noise sources in the spatial perception experiment. The distance is defined as the space between the listener position and the center of the circle placed by the listener. The actual loudspeaker distance of 2.4 m is indicated by the gray horizontal line. The boxplots indicate the median and the first and third quartiles. The whiskers extend to 1.5 times the interquartile range. Outliers are indicated as dots.

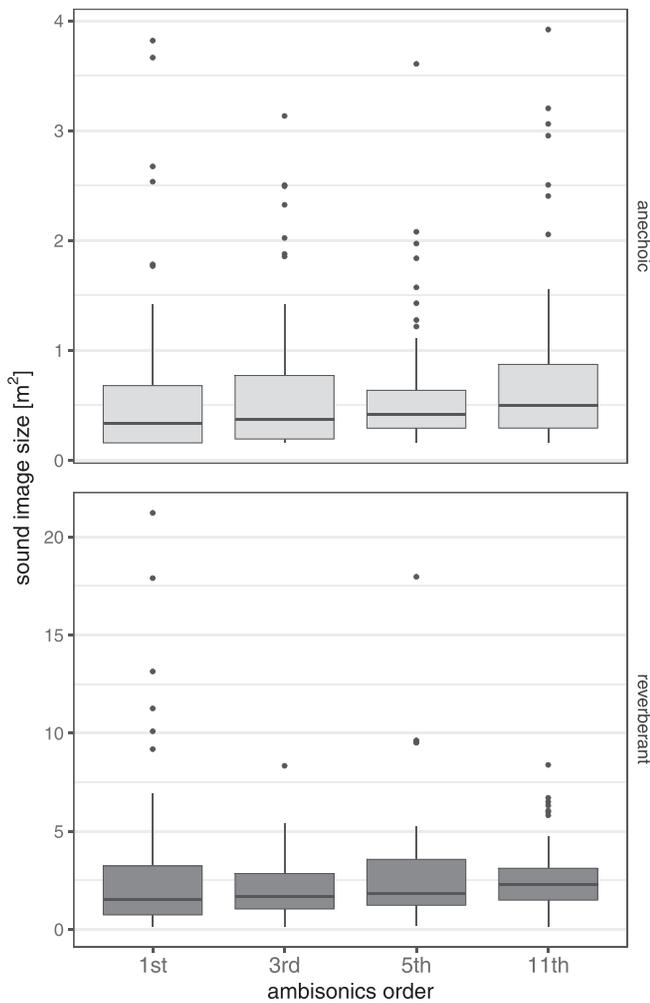


FIG. 5. Sound image size in the spatial perception experiment, calculated as the area of the reported source circle in the anechoic (top) and reverberant (bottom) conditions. The boxplots indicate the median and the first and third quartiles. The whiskers extend to 1.5 times the interquartile range. Outliers are indicated as dots.

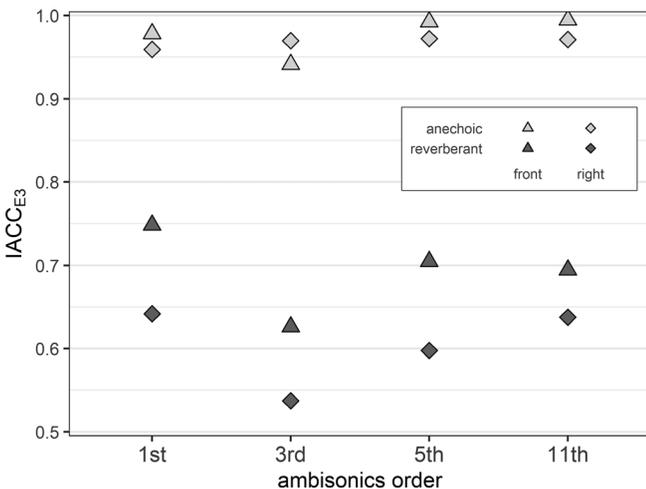


FIG. 6. The IACC coefficient of the first 80 ms of the impulse responses averaged over three octave bands (0.5, 1, and 2 kHz; $IACC_{E3}$) as a function of the ambisonics order. The $IACC_{E3}$ is shown for the anechoic (light gray) and reverberant (dark gray) conditions for sources from the front (0° , on-axis, triangles) and from 15° to the right (diamonds).

width. Considering the results of the current experiment, as well as those of Bertet *et al.* (2013) and Whitmer *et al.* (2014), low-order ambisonics processing may distort spatial cues in a way that affects localization but not the sound image size perception.

While no effect of the ambisonics order on the sound image size was found, the results of the present experiment did show an effect of ambisonics order on the perceived distance. In the anechoic condition, listeners perceived the stimuli presented with low ambisonics order to be closer to them than the higher-order stimuli, while in the reverberant condition no differences were found. Without reverberation, the direct-to-reverberant ratio, which is a major cue for distance perception, is not available (Zahorik *et al.*, 2005). Thus, in the absence of alternate distance cues, listeners might have interpreted the wider spread of energy as a cue for the perceived distance instead of the size. In the reverberant conditions, the listeners perceived the sources at the correct distance. In addition to purely auditory cues, the incongruence between the auditory stimuli (representing a small reverberant room) and visual stimulus (a large anechoic chamber) may have made the subjective judgments of distance more difficult (Gil-Carvajal *et al.*, 2016).

IV. EXPERIMENT 2: SPEECH INTELLIGIBILITY WITH TWO INTERFERING TALKERS FIXED IN SPACE

A. Methods

Experiment 2 investigated the influence of the spatial energy spread on speech intelligibility. The speech material of the target and two interfering talkers was taken from the multi-talker version of the Danish matrix sentence test Dantale II. Dantale II sentences have a name-verb-numeral-adjective-noun structure. The name was presented as a call sign, and the listeners were asked to identify the remaining four words on a UI displayed on the same touchscreen as in experiment 1. The call sign was continuously shown on the UI. For each word category, ten words exist in the speech test and are shown as possible response alternatives. The responses were scored on a word basis, and speech reception thresholds (SRTs) were estimated with an adaptive procedure by varying the target-to-masker ratio (TMR), converging at 70% correct intelligibility (Brand and Kollmeier, 2002). The adaptive procedure was terminated after 8 reversals if at least 20 sentences had been presented. The SRTs were calculated as the average TMR of the last six reversals. The sound pressure level (SPL) of each masker was kept constant at 60 dB, while the level of the target speech was adjusted adaptively, starting at 70 dB. The speech material contained five female talkers with a similar voice pitch. However, only three talkers (talkers 1, 4, and 5) were chosen because the average level of the other two talkers differed strongly.

SRTs were measured in two spatial configurations: a colocated condition with the target and two interfering talkers presented from the front (0° , on-axis) and a separated condition with the target from the front and the two

interferers presented from $\pm 15^\circ$ azimuth. For each SRT measurement, a call sign (name) was chosen randomly and kept for all sentences, whereas the three talkers representing the target and interfering sources were chosen randomly for each sentence.

Each listener was introduced to and familiarized with the task by presenting 5–10 sentences in quiet. SRTs were then measured in the conditions with the different ambisonics orders, with and without reverberation, and with colocated and separated interferers, leading to 16 ($4 \times 2 \times 2$) SRT measurements overall. The conditions were presented in random order to the listeners.

B. Results and discussion

Figure 7 shows results obtained in the speech intelligibility experiment for the anechoic condition (top panel) and the reverberant condition (bottom panel) with spatially colocated (white boxes) and separated (gray boxes) interferers. The statistical analysis of the SRTs revealed significant main effects [order, $F(3,186) = 10.8, p < 0.0001$; interferer configuration, $F(1,186) = 321.8, p < 0.0001$; reverberation,

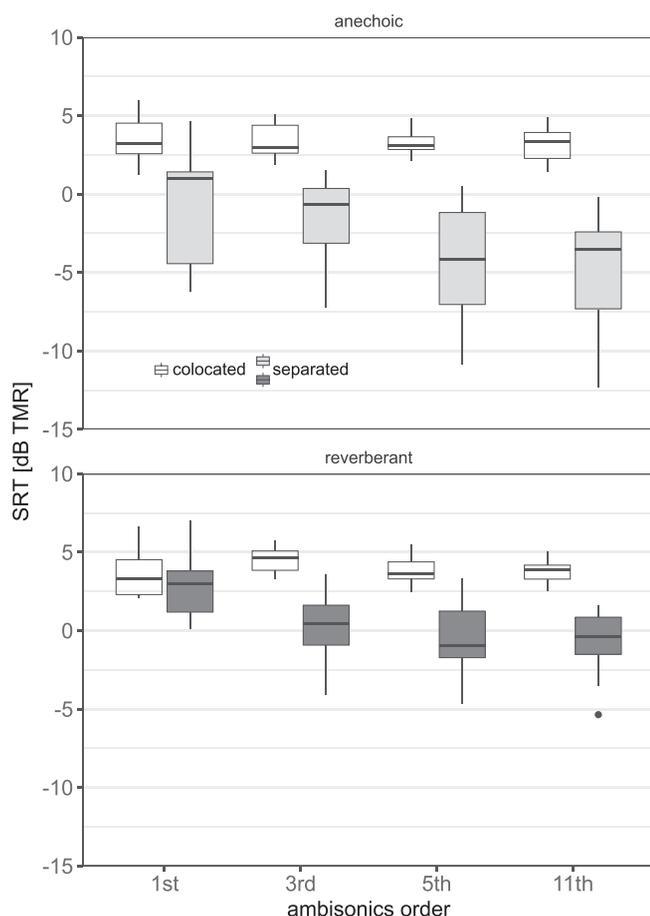


FIG. 7. SRTs at 70% correct as target-to-masker ratio in dB with two colocated (white boxplots) and two symmetrically separated interferers (gray boxplots). The top panel represents the anechoic condition and the bottom panel represents the reverberant condition. The boxplots indicate the median and the first and third quartiles. The whiskers extend to 1.5 times the interquartile range. Outliers are indicated as dots.

$F(1,186) = 51.4, p < 0.0001$], as well as significant interactions between ambisonics order and interferer configuration [$F(3,186) = 10.1, p < 0.0001$] and reverberation and interferer configuration [$F(1,186) = 22.1, p < 0.0001$].

In the colocated interferer configuration, no differences were found between any of the ambisonics orders ($p = 1$). Similarly, no effect of reverberation was found when the target and interfering talkers were colocated [$t(186) = -1.7, p = 0.17$]. These findings are consistent with previous work with this speech material (Ahrens *et al.*, 2019). It has been argued that a positive TMR is needed to segregate the sources in situations with similar target and interfering speech material and no spatial separation (Brungart *et al.*, 2001; Best *et al.*, 2012), which might obscure effects of ambisonics order and reverberation.

Further analysis was performed on the SRM, the difference between the colocated and the separated interferer configurations. Figure 8 shows the SRM obtained in the anechoic (light gray boxes) and reverberant (dark gray boxes) conditions as a function of the ambisonics order. The analysis of the linear mixed model with the ambisonics order and reverberation condition as fixed effects and the listeners as a random effect revealed significant contributions of both main effects [order, $F(3,87) = 12.4, p < 0.0001$; reverberation, $F(1,87) = 27.1, p < 0.0001$] but no interaction [$F(3,84) = 1.7, p = 0.17$]. The *post hoc* analysis between the orders revealed that the SRM in the 1st-order ambisonics condition was smaller than for the higher ambisonics orders [3rd, $t(87) = -2.9, p = 0.0268$; 5th, $t(87) = -4.9, p < 0.0001$; 11th, $t(87) = -5.5, p < 0.0001$]. The differences between the 3rd and 5th order [$t(87) = -2.0, p = 0.3159$] and between the 5th and 11th order [$t(87) = -0.7, p = 1$] were not found to be significant. Even though the difference between the 3rd and 11th orders was slightly above the traditional significance level of 0.05 [$t(87) = -2.6, p = 0.0617$], after Bonferroni correction, a trend of an increase of the SRM with increasing

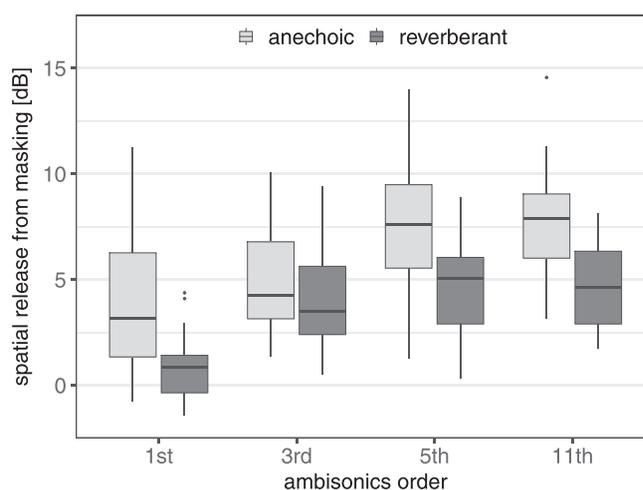


FIG. 8. The measured SRM (boxplots) in the anechoic (light gray) and reverberant (dark gray) conditions. The boxplots indicate the median and the first and third quartiles. The whiskers extend to 1.5 times the interquartile range. Outliers are indicated as dots.

ambisonics order was found. Therefore, the ambisonics presentation order and, thus, the spatial energy spread affected the SRM.

However, it is not clear whether the reduced SRM at low orders is related to the spatial position of the source, i.e., whether speech intelligibility can be restored by increasing the source-target separation. This was considered in the following experiment, where the target-masker separation angle was investigated.

V. EXPERIMENT 3: SPEECH INTELLIGIBILITY WITH TWO INTERFERING TALKERS VARYING IN SPACE

A. Methods

Experiment 3 investigated speech intelligibility of target sentences from the front (0°, on-axis) in the presence of spatially varying interfering talkers. This was done for a fixed TMR of -6 dB in the anechoic condition and for the same ambisonics orders (1st, 3rd, 5th, and 11th order) corresponding to different degrees of energy spread as described above. With the fixed TMR, the separation angle of two symmetrically separated interferers was varied to obtain 70% speech intelligibility (speech reception threshold angle, SRA). The adaptive procedure was terminated after 8 reversals if at least 25 sentences had been presented. The SRAs were calculated as the average separation angle of the last six reversals. The particular TMR was chosen based on pilot testing and set to obtain a reasonable range of angles, avoiding ceiling and floor effects. The speech material was the same as in experiment 2, where the interferers had fixed spatial locations. The SRA was measured using an adaptive procedure as described in Brand and Kollmeier (2002). The separation angle of a specific trial was calculated using the same procedure as was used to obtain the SRT in experiment 2 (Brand and Kollmeier, 2002). The change in separation angle ($\Delta\Theta$) of the subsequent trial was defined as

$$\Delta\Theta = -\frac{f(n)(\text{prev} - \text{tar})}{\text{slope}},$$

where n is the reversal number, “prev” refers to the discrimination value of the previous sentence, and “tar” refers to the discrimination value to which the procedure converges.

The parameters $f(n)$ and slope were adapted from the recommendations provided by Brand and Kollmeier (2002) to account for the fact that the separation angle was used here as a tracking variable instead of the TMR. This was needed to adjust for the different numerical ranges of the TMR and the separation angle. A slope parameter of 0.029 deg^{-1} and an $f(n) = 1.5 \times 1.15^{-n}$ were used to obtain the different step sizes.

The range of separation angles was limited to where speech intelligibility was expected to be a monotonic function of separation angle. The minimum angle was set to 0°, since the highest SRT has commonly been found at 0° separation (i.e., colocated). The maximum separation angle was set to $\pm 105^\circ$ to cover a wide range of angles, well above the

angle of $\pm 45^\circ$ that has previously been shown to lead to the lowest SRT (Marrone *et al.*, 2008). The initial separation angle between the target and interferers was 75°. Each listener repeated each condition twice. The repetitions were treated as a fixed effect to investigate a possible training effect.

B. Results and discussion

Figure 9 shows the angle between the target and the two symmetric interferers that is needed to identify 70% of the words correct (SRA). The statistical analysis of the SRA revealed a significant effect of the ambisonics order [$F(3,67) = 11.6, p < 0.0001$] but not of the repetitions [$F(1,66) = 1.3, p = 0.26$] and their interaction [$F(3,63) = 1.2, p = 0.32$]. Thus, no training effect over the repetitions was found. The result of the *post hoc* analysis is shown in Table I. Generally, smaller SRAs were found for the higher ambisonics orders. However, when comparing 1st vs 3rd order, as well as 5th vs 11th order, no significant differences were found.

Comparing the variances of the SRA across ambisonics orders, it is apparent that the variance for the third-order responses is larger than for the other conditions. As it can be seen from Fig. 9, some listeners performed comparably to the 5th/11th-order conditions, while some listeners performed more similarly to the 1st-order condition, or even obtained higher SRAs than for the 1st order. However, it is unclear why in this particular condition listeners behaved this way. It is possible that while most listeners were able to utilize the more detailed spatial cues at the higher orders, at third order only some listeners were able to take advantage of the additional information compared to the first order.

There is a potential risk that speech intelligibility may not have been a monotonic function with respect to the

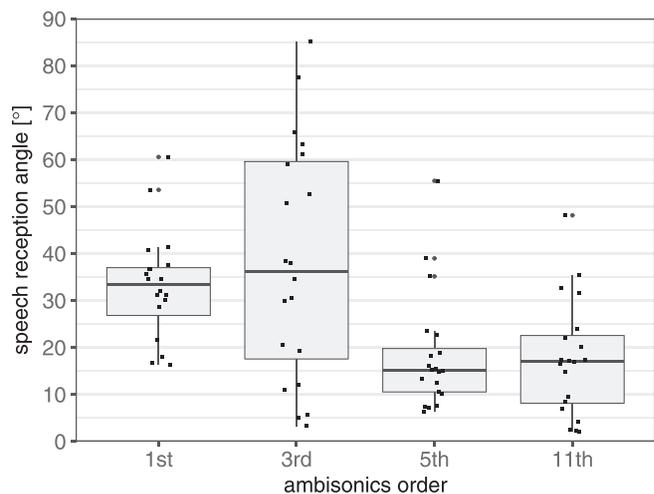


FIG. 9. SRA, i.e., the separation angle between the target and two symmetrically spaced interferers that leads to 70% intelligibility, at -6 dB target-to-masker ratio in the anechoic condition. The boxplots indicate the median and the first and third quartiles. The whiskers extend to 1.5 times the interquartile range. Outliers are indicated as dots. The black squares are single listeners' responses.

TABLE I. Results from the *post hoc* analysis of the speech reception threshold angle (SRA). Nonsignificant results with *p*-value larger than 0.05 are indicated in gray.

	1st order	3rd order	5th order	11th order
1st order	–	$t(67) = -1.2$ $p = 1.0$	$t(67) = 3.4$ $p = 0.0074$	$t(67) = 3.6$ $p = 0.0043$
3rd order		–	$t(67) = 4.6$ $p = 0.0001$	$t(67) = 4.8$ $p = 0.0001$
5th order			–	$t(67) = 0.2$ $p = 1.0$
11th order				–

separation angle, which could have led to a non-converging behavior in the adaptive procedure. However, no anomalies were observed in the adaptive tracks or the reconstructed psychometric functions. The corresponding data are provided in the supplemental material.¹

The results show that the changes in speech intelligibility due to the varying energy spread do relate to the spatial position of the sources: Sources with a larger energy spread require a larger angular separation for equal intelligibility when comparing the 1st and 11th orders. The general size of the SRA is consistent with results from Lócsei *et al.* (2017), who measured the interaural time difference needed to understand 50% of the words to produce a SRM of 3 dB measured with a two-talker babble noise. Their results varied between 140 and 370 μ s, which corresponds to about 15°–45° azimuth location as measured on an artificial head (e.g., Oreinos and Buchholz, 2013) or estimated from a head model (Aaronsen and Hartmann, 2014). These angles are above the SRA found in the current study for sources reproduced with higher ambisonics orders (low energy spread), which can be explained by the lower criterion used in Lócsei *et al.* (2017; 3 dB vs 6 dB in the current study).

VI. OVERALL DISCUSSION AND SUMMARY

In the present study, three experiments were conducted to investigate the effect of the spatial spread of energy on speech perception in young normal-hearing listeners. In experiment 1, it was shown that a wider energy spread elicited by ambisonics processing did not lead to perceptually larger sound images. Correspondingly, the IACC_{E3}, a physical correlate for the apparent source size, was not found to vary with the energy spread either. Instead, sources were perceived as being closer in distance but only when presented anechoically. In experiment 2, a lower SRM was found for sources with a large energy spread than for sources with a low energy spread. In the third experiment, the minimum separation angle between a target speech and interfering speech sources in terms of speech intelligibility was found to be related to the energy spread. For equal speech intelligibility, a wider separation was needed for sources with a large energy spread than for sources with a low energy spread. However, the trend was inconsistent and a large variance across listeners was found. The results from

experiments 2 and 3 also suggest that 5th-order ambisonics may be adequate as a reproduction method for a speech intelligibility task with a 15° target-masker separation, as no differences were found vs the higher, 11th-order conditions.

The aim of this study was to investigate a possible connection between spatial energy spread, sound image size, and speech intelligibility. The results showed that the energy spread affected speech intelligibility but not the sound image size. The percept of sound image size has previously been related to binaural features such as the IACC_{E3} or fluctuations of interaural time differences (Griesinger, 1997; Mason *et al.*, 2001; Whitmer *et al.*, 2012). In the current study, the IACC_{E3} was considered and found not to vary systematically with the spatial energy spread, which is in agreement with the finding of no differences in the perceptual estimates of size. This suggests that a large energy spread may not be a good indicator of sound sources that are typically perceived by young normal-hearing listeners as large, such as ensembles of similar sources, or sources with a large physical extent. Listeners may have also had difficulties in labeling the sizes of the sound images they perceived as the concept of size, particularly that of a voice, might not be natural or obvious for untrained listeners. Nevertheless, previous studies investigating the sound image percept have shown that normal-hearing listeners are able to assign a size to speech stimuli (e.g., Hassager *et al.*, 2017; Cubick *et al.*, 2018).

Additionally, varying the ambisonics order does not only control the energy spread but also introduces varying magnitude and phase errors at higher frequencies due to different frequency range limitations for different orders (Daniel, 2001). While equalization and dual-band decoding were used to reduce these errors, the sound field at the ear positions of the listeners may have differed in other aspects than purely the energy spread of the sources. This, in turn, may have resulted in speech intelligibility degradations that were not related to the spatial energy spread. While such contributions cannot be excluded, the results from experiment 3 demonstrated a small but significant effect of ambisonics order on spatial separation. Errors in sound pressure at the listener’s head with ambisonics reproduction are most prominent at the ear contralateral to the sound source, with low-order ambisonics effectively reducing the available head-shadow advantage at higher frequencies (Oreinos and Buchholz, 2015). The larger separation angle (experiment 3) and lower SRM (experiment 2) found for low ambisonics orders may have been a consequence of this reduced head-shadow advantage.

Thus, the disconnect between the perceived sound image size and speech intelligibility may be caused by different underlying cues. The percept of sound image size has been linked to the IACC, which did not change with ambisonics order in a systematic way. Speech intelligibility, on the other hand, depends on the TMR at the ears (Zurek, 1993; Glyde *et al.*, 2013), as well as any binaural interactions (Durlach, 1963, 1972; Culling *et al.*, 2004), which may have been affected by the ambisonics processing. This implies that any processing that influences the spatial spread

of energy, for example through a low-order reproduction in ambisonics-based virtual sound environments, can lead to degraded speech intelligibility even when the perception of spatial extent is unaffected. Therefore, speech tests utilizing such sound reproduction methods may need to consider whether sound sources are reproduced with a sufficiently low energy spread in relation to the minimum expected source separation.

ACKNOWLEDGMENTS

This work was supported by the Technical University of Denmark and the Oticon Centre of Excellence for Hearing and Speech Sciences (CHeSS). The multi-talker version of the Dantale II speech material was provided by Eriksholm Research Centre. We would like to thank Adam Westermann for an early version of the graphical user test interface for the speech test and Henrik Hassager for the graphical UI for the spatial perception experiment, as well as Johannes Käsbach for the fruitful discussions regarding source width perception. We would also like to thank the editor, Virginia Best, as well as the two anonymous reviewers, for their valuable comments.

¹See supplementary material at <https://zenodo.org/record/3497462> for data from the three experiments.

- Aaronson, N. L., and Hartmann, W. M. (2014). "Testing, correcting, and extending the Woodworth model for interaural time difference," *J. Acoust. Soc. Am.* **135**, 817–823.
- Ahrens, A. (2018). "Room acoustics model of a listening room [dataset]," available at <https://doi.org/10.5281/zenodo.1232317> (Last viewed 2/10/2020).
- Ahrens, A., Marschall, M., and Dau, T. (2019). "Measuring and modeling speech intelligibility in real and loudspeaker-based virtual sound environments," *Hear. Res.* **377**, 307–317.
- Behrens, T., Neher, T., and Johannesson, B. (2007). "Evaluation of a Danish speech corpus for assessment of spatial unmasking," in *Auditory Signal Processing in Hearing-Impaired Listeners, 1st International Symposium on Auditory and Audiological Research (ISAAR 2007)*, pp. 449–457.
- Bertet, S., Daniel, J., Gros, L., and Parizet, E. (2007). "Investigation of the perceived spatial resolution of higher order ambisonics sound fields: A subjective evaluation involving virtual and real 3D microphones," in *AES 30th International Conference*, Saariselkä, Finland (15–17 March 2007) pp. 1–9.
- Bertet, S., Daniel, J., Parizet, E., and Warusfel, O. (2013). "Investigation on localisation accuracy for first and higher order ambisonics reproduced sound sources," *Acta Acust. Acust.* **99**, 642–657.
- Best, V., Marrone, N., Mason, C. R., and Kidd, G., Jr. (2012). "The influence of non-spatial factors on measures of spatial release from masking," *J. Acoust. Soc. Am.* **131**, 3103–3110.
- Best, V., Thompson, E. R., Mason, C. R., and Kidd, G. (2013). "An energetic limit on spatial release from masking," *J. Assoc. Res. Otolaryngol.* **14**, 603–610.
- Blauert, J. (1984). *Spatial Hearing: The Psychophysics of Human Sound Localization* (MIT Press, Cambridge, MA).
- Blauert, J., and Lindemann, W. (1986a). "Auditory spaciousness: Some further psychoacoustic analyses," *J. Acoust. Soc. Am.* **80**, 533–542.
- Blauert, J., and Lindemann, W. (1986b). "Spatial mapping of intracranial auditory events for various degrees of interaural coherence," *J. Acoust. Soc. Am.* **79**, 806–813.
- Brand, T., and Kollmeier, B. (2002). "Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests," *J. Acoust. Soc. Am.* **111**, 2801–2810.
- Brungart, D. S., Simpson, B. D., Ericson, M. A., and Scott, K. R. (2001). "Informational and energetic masking effects in the perception of multiple simultaneous talkers," *J. Acoust. Soc. Am.* **110**, 2527–2538.
- Cubick, J., Buchholz, J. M., Best, V., Lavandier, M., and Dau, T. (2018). "Listening through hearing aids affects spatial perception and speech intelligibility in normal-hearing listeners," *J. Acoust. Soc. Am.* **144**, 2896–2905.
- Culling, J. F., Hawley, M. L., and Litovsky, R. Y. (2004). "The role of head-induced interaural time and level differences in the speech reception threshold for multiple interfering sound sources," *J. Acoust. Soc. Am.* **116**, 1057–1065.
- Daniel, J. (2001). "Représentation de champs acoustiques, application à la transmission et à la reproduction de scènes sonores complexes dans un contexte multimédia" ("Acoustic field representation, application to the transmission and the reproduction of complex sound environments in a multimedia context"), Ph.D. thesis, Université Paris 6.
- Duquesnoy, A. J. (1983). "Effect of a single interfering noise or speech source upon the binaural sentence intelligibility of aged persons," *J. Acoust. Soc. Am.* **74**, 739–743.
- Durlach, N. I. (1963). "Equalization and cancellation theory of binaural masking-level differences," *J. Acoust. Soc. Am.* **35**, 1206–1218.
- Durlach, N. I. (1972). "Binaural signal detection equalization and cancellation theory," in *Binaural Signal Detection Equalization and Cancellation Theory* (Academic Press, New York), pp. 371–462.
- Favrot, S., and Buchholz, J. M. (2010). "LoRA: A loudspeaker-based room auralization system," *Acta Acust. Acust.* **96**, 364–375.
- Frank, M. (2013). "Source width of frontal phantom sources: Perception, measurement, and modeling," *Arch. Acoust.* **38**, 311–319.
- Gerzon, M. (1973). "Periphony: With-height sound reproduction," *J. Audio Eng. Soc.* **21**, 2–10.
- Gerzon, M. (1992). "General metatheory of auditory localisation," in *92nd Convention of Audio Engineering Society*, Vienna.
- Gil-Carvajal, J. C., Cubick, J., Santurette, S., and Dau, T. (2016). "Spatial hearing with incongruent visual or auditory room cues," *Sci. Rep.* **6**, 37342.
- Glyde, H., Buchholz, J. M., and Dillon, H. (2013). "The importance of interaural time differences and level differences in spatial release from masking," *J. Acoust. Soc. Am.* **134**, 2169–2180.
- Griesinger, D. (1997). "The psychoacoustics of apparent source width, spaciousness and envelopment in performance spaces," *Acta Acust.* **83**, 721–731.
- Hassager, H. G., Wiinberg, A., and Dau, T. (2017). "Effects of hearing-aid dynamic range compression on spatial perception in a reverberant environment," *J. Acoust. Soc. Am.* **141**, 2556–2568.
- Hering, E. (1861). *Beitraege zur Physiologie* (Contributions to physiology) (Engelmann, Leipzig).
- Holway, A. H., and Boring, E. G. (1941). "Determinants of apparent visual size with distance variant," *Am. J. Psychol.* **54**, 21–37.
- IEC 268-13 (1985). *Sound System Equipment Part 13: Listening Tests on Loudspeaker* (International Electrotechnical Commission, Geneva, Switzerland).
- Kuznetsova, A., Christensen, R. H. B., Bavay, C., and Brockhoff, P. B. (2014). "Automated mixed ANOVA modeling of sensory and consumer data," *Food Qual. Prefer.* **40**, 31–38.
- Lenth, R. V. (2016). "Least-squares means: The R package lsmeans," *J. Stat. Softw.* **69**, 1–33.
- Löcsei, G., Santurette, S., Dau, T., and Macdonald, E. (2017). "Lateralized speech perception with small ITDs in normal-hearing and hearing-impaired listeners," in *Proceedings of the International Symposium on Auditory and Audiological Research (Proc. ISAAR): Adaptive Processes in Hearing*.
- Marrone, N., Mason, C. R., and Kidd, G. (2008). "Tuning in the spatial dimension: Evidence from a masked speech identification task," *J. Acoust. Soc. Am.* **124**, 1146–1158.
- Martin, R. L., McAnally, K. I., Bolia, R. S., Eberle, G., and Brungart, D. S. (2012). "Spatial release from speech-on-speech masking in the median sagittal plane," *J. Acoust. Soc. Am.* **131**, 378–385.
- Mason, R., Rumsey, F., and De Bruyn, B. (2001). "An investigation of interaural time difference fluctuations, Part I: The subjective spatial effect of fluctuations delivered over headphones," in *110th AES Convention*, Amsterdam.

- Okano, T., Beranek, L. L., and Hidaka, T. (1998). "Relations among interaural cross-correlation coefficient (IACCE), lateral fraction (LFE), and apparent source width (ASW) in concert halls," *J. Acoust. Soc. Am.* **104**, 255–265.
- Oreinos, C., and Buchholz, J. M. (2013). "Measurement of a full 3D set of HRTFs for in-ear and hearing aid microphones on a head and torso simulator (HATS)," *Acta Acust. Acust.* **99**, 836–844.
- Oreinos, C., and Buchholz, J. M. (2015). "Objective analysis of ambisonics for hearing aid applications: Effect of listener's head, room reverberation, and directional microphones," *J. Acoust. Soc. Am.* **137**, 3447–3465.
- Oreinos, C., and Buchholz, J. M. (2016). "Evaluation of loudspeaker-based virtual sound environments for testing directional hearing aids," *J. Am. Acad. Audiol.* **27**, 541–556.
- Politis, A. (2016). "Microphone array processing for parametric spatial audio techniques," Ph.D. thesis, Aalto University.
- Solvang, A. (2008). "Spectral impairment for two-dimensional higher order ambisonics," *AES J. Audio Eng. Soc.* **56**, 267–279.
- Wagener, K., Jøsvassen, J. L., and Ardenkjær, R. (2003). "Design, optimization and evaluation of a Danish sentence test in noise," *Int. J. Audiol.* **42**, 10–17.
- Westermann, A., and Buchholz, J. M. (2015). "The effect of spatial separation in distance on the intelligibility of speech in rooms," *J. Acoust. Soc. Am.* **137**, 757–767.
- Whitmer, W. M., Seeber, B. U., and Akeroyd, M. A. (2012). "Apparent auditory source width insensitivity in older hearing-impaired individuals," *J. Acoust. Soc. Am.* **132**, 369–379.
- Whitmer, W. M., Seeber, B. U., and Akeroyd, M. A. (2014). "The perception of apparent auditory source width in hearing-impaired adults," *J. Acoust. Soc. Am.* **135**, 3548–3559.
- Wiggins, I. M., and Seeber, B. U. (2011). "Dynamic-range compression affects the lateral position of sounds," *J. Acoust. Soc. Am.* **130**, 3939–3953.
- Wiggins, I. M., and Seeber, B. U. (2012). "Effects of dynamic-range compression on the spatial attributes of sounds in normal-hearing listeners," *Ear Hear.* **33**, 399–410.
- Zahorik, P., Brungart, D. S., and Bronkhorst, A. W. (2005). "Auditory distance perception in humans: A summary of past and present research," *Acta Acust. Acust.* **91**, 409–420.
- Zotter, F., and Frank, M. (2012). "All-round ambisonic panning and decoding," *AES J. Audio Eng. Soc.* **60**, 807–820.
- Zotter, F., Frank, M., Kronlacher, M., and Choi, J.-W. (2014). "Efficient phantom source widening and diffuseness in ambisonics," in *Proceedings of the EAA Joint Symposium on Auralization and Ambisonics*, Berlin, Germany (3–5 April 2014) Vol. 2, pp. 69–74.
- Zurek, P. M. (1993). "Binaural advantages and directional effects in speech intelligibility," in *Acoust. Factors Affect. Hear. Aid Perform.* **2**, 255–175.