The Journal of the Acoustical Society of America

OCTOBER 04 2024

Effect of ambisonic order on spatial release from masking

John F. Culling ; A. Ottopea Akrofi; Samuel R. Dighton

J. Acoust. Soc. Am. 156, 2169–2176 (2024)
https://doi.org/10.1121/10.0030462

ASA ACOUSTICAL SOCIETY OF AMERICA

JASA ARTICLE

Effect of ambisonic order on spatial release from masking

John F. Culling,[a] A. Ottopea Akrofi, and Samuel R. Dighton
School of Psychology, Cardiff University, 70 Park Place, Cardiff CF10 3AT, United Kingdom

ABSTRACT:
In two experiments, spatial release from masking (SRM) was measured using an ambisonic reproduction system for a range of different ambisonic orders. The first experiment used ambisonic panning while the second experiment used impulse responses recorded from a sixth-order ambisonic microphone. Both experiments found a progressive increase in SRM with increasing ambisonic order for speech presented against a single speech-shaped-noise interferer. SRM increased progressively up to at least fourth order and continued to asymptotically improve at higher-ambisonic orders toward the level achieved with point sources. The second experiment found that this effect was robust for different acoustic environments.
 https://doi.org/10.1121/10.0030462

(Received 29 April 2024; revised 6 September 2024; accepted 16 September 2024; published online 4 October 2024)

[Editor: Jonas Braasch] Pages: 2169–2176

I. INTRODUCTION

Ambisonics is a method for reproducing a sound field that has been recorded at a particular location using multiple directional microphones (Gerzon, 1973). The raw recording is encoded into a format that encodes the spatial distribution of incoming sound into spherical-harmonic channels. This *B*-format signal can then be decoded into loudspeaker channels for reproduction over a specified spatial configuration of loudspeakers. The array of loudspeakers is typically distributed over the surface of a sphere or just a circle in the horizontal plane with the listener located at the center. The quality of reproduction is best within a "sweet-spot" area around this central position. Until recently, the demands of creating such a reproduction arrangement have limited its use to specialist listening spaces and research facilities, but virtual reality applications delivered over headphones (Noisternig *et al.*, 2003) have now made it a widespread consumer product.

Most ambisonic systems are first order (Zotter and Frank, 2019). Here, recordings, which are ultimately reproduced over at least four loudspeakers, are made by four directional microphones in a tetrahedral arrangement. However, in recent years, a number of systems have been developed that extend the spherical-harmonic order. In principle, higher-order ambisonics should increase the spatial resolution of the reproduced sound at the expense of deploying more microphones, more signal processing, and more loudspeakers, but the perceptual benefits of this investment have not been extensively explored. Bertet *et al.* (2013), Stitt *et al.* (2014), and Huisman *et al.* (2021) have shown that sound localization is improved using higher-order ambisonic microphones compared to that achieved with a first-order microphone, whereas McCormack *et al.* (2020) have shown a progressive improvement across ambisonic orders 1, 3, and 5 in overall evaluations of reproduction quality for musical instrument sounds in simulated auditoria. The present study examines the effect of ambisonic order on spatial release from masking (SRM), the benefit of introducing an angular separation between speech and interfering noise on speech intelligibility. We are aware of three previous studies that have addressed the same question, but their methodologies differed in many ways, and the results appear somewhat inconsistent.

Dagan *et al.* (2019) found a SRM of approximately 5 dB using virtual ambisonic presentation for digit triplets and noise at ±45°. The baseline condition for the measurement of SRM was an ambisonic order of zero, which encodes no spatial information. Virtual ambisonics is a technique in which the loudspeaker system is simulated over headphones using head-related transfer functions (HRTFs) between each loudspeaker and the listener's head (Noisternig *et al.*, 2003). The technique is widely used in virtual reality because dynamic update of the stimuli for rotation of the head can be achieved through computationally efficient rotation in the spherical-harmonic domain or by a dynamic update of the HRTFs for a fixed set of virtual loudspeakers. In the experiment by Dagan *et al.* (2019), no head-tracking was employed, therefore, the listeners were effectively listening with a fixed head position. The signals to each virtual loudspeaker were also based on prediction rather than recording from a microphone array. The results showed that SRM did not improve for ambisonic orders from 1 to 32, suggesting that first-order ambisonics may be sufficient for the full effect of SRM, and the use of higher orders is unnecessary. However, the maximum SRM observed fell substantially short of what would be expected from theory for point sources and a continuo us noise

[a]Email: cullingj@cf.ac.uk

masker; the Jelfs model (Jelfs *et al.*, 2011) predicts a SRM of 12 dB in anechoic conditions for this stimulus configuration using HRTFs from Gardner and Martin (1995).

Ahrens *et al.* (2019) demonstrated that a seventh-order ambisonic end-to-end recording and reproduction system can recreate the sound field from a moderately reverberant listening room ($T_{30} = 400$ ms) with sufficient accuracy to produce similar observed SRM to direct presentation of the stimuli in that room. The spatialized configurations placed a target matrix sentence (Dantale II, Wagener *et al.* 2009) at 0° and two interfering talkers or speech-modulated noises either at ±30° or both at 30° while the baseline condition had all sources co-located at 0°. Stimuli were recorded using a bespoke 52-channel spherical microphone and reproduced in an anechoic chamber over a spherical array of 64 loudspeakers. Because only a single ambisonic order was employed, it is unclear how the quality of reproduction increases as a function of ambisonic order and, therefore, whether first order might, again, have been sufficient. Benefits of spatial separation with ambisonic reproduction were limited to cases with speech interferers which could produce a degree of informational masking (Brungart, 2001), making any comparison with SRM that has been predicted based on energetic masking (e.g., Jelfs *et al.*, 2011) unreliable.

In a later study in which different ambisonic orders were rendered on a circular loudspeaker array, Ahrens *et al.* (2020) found that SRM increased with increasing ambisonic order, but the only statistically significant differences were between 1st order and orders 3, 5, or 11. The increase in SRM between 3rd and 5th order was nonsignificant, and 5th and 11th order gave similar levels of SRM. The contrast between first order and each of the higher orders disagrees with the conclusion by Dagan *et al.* (2019) that first order is sufficient. However, the spatial distribution of the sound sources was narrower than either of the previous studies with the target speech at 0° and interfering voices at ±15°. Consequently, it is unclear whether the benefit of increased spatial resolution that higher-order ambisonics can convey is limited to sound sources that are more closely spaced or situations in which information masking is involved. Other details of the experiment were also unique. Ambisonic encoding was simulated using a room auralization toolbox (Favrot and Buchholz, 2009) and decoded onto a circular array of loudspeakers in an anechoic chamber. By using this synthetic technique and limiting the ambisonic decoding to the horizontal plane, a higher order of ambisonic reproduction was possible. The software simulated their reverberant room from the previous study and also an anechoic environment.

Overall, there are many potential methodological options to choose from for asking the same empirical question. Beyond the differences highlighted above, it is also possible to decode the signals onto the loudspeakers using different algorithms. In principle, all of these options should be equivalent and produce similar results, but the work of Dagan *et al.* (2019) and Ahrens *et al.* (2020) led to inconsistent conclusions. The present study provides additional evidence using two new combinations of methods. In experiment 1, ambisonic panning (Neukom, 2007) is used to present speech against interfering noise on a circular array of loudspeakers at a range of ambisonic orders and for a range of spatial separations. Importantly, it includes an explicit comparison with point-source presentation for each spatial configuration such that the asymptotic approach to point-source performance can be assessed. In experiment 2, a sixth-order ambisonic microphone was used to test the same question using an end-to-end recording and reproduction system. Impulse responses were recorded in two rooms, one reverberant and one dry. The impulse responses were encoded at a range of ambisonic orders and convolved with speech and noise for presentation over the same loudspeaker array. Consistent with Ahrens *et al.* (2020), both experiments found a progressive increase in SRM with ambisonic order, but with the improvements being statistically significant up to and beyond order 4.

II. EXPERIMENT 1

A. Equipment

The stimuli were presented using a 48-loudspeaker circular array of 1.2 m radius, located in a sound-deadened laboratory with sound absorbing wall and ceiling panels and a carpeted floor ($T60 \approx 60$ ms). The loudspeakers were Minx 10, satellite loudspeakers (Cambridge Audio, London), driven by 8, 6-channel car amplifiers (Auna, Berlin) and two 24-channel audio interfaces (Motu 24I/O Core, Cambridge, MA).

B. Stimuli

Speech targets were IEEE (Institute of Electrical and Electronic Engineers) sentences (Rothauser *et al.*, 1969) from the Harvard recordings (male voice DA) and interfering noise was filtered to match the long-term excitation pattern (Moore and Glasberg, 1983) of all the speech material. The IEEE sentences are semantically plausible but not very predictable. They come in lists of ten and each contains five keywords. One example, with keywords in capitals, is "GLUE the SHEET to the DARK BLUE BACKGROUND."

Five seconds of Gaussian noise was used as a masker, which exceeded the duration of the longest speech stimulus. The noise was filtered with a 512-point finite-impulse-response filter to provide speech-shaping and then gated with 100-ms raised-cosine onset and offset ramps.

The speech and noise were rendered onto the array of 48 loudspeakers using the 'basic' method of ambisonic panning (Neukom, 2007). For mth-order panning with n loudspeakers in a circle, a weight, f, (either positive or negative) is applied to the signal for the loudspeaker at azimuth Θ. Equation (1) is reproduced from Neukom (2007). Equation (2) adapts their equation to always render ambisonic order m with $2m+1$ loudspeakers, which is the minimum number required on a circular loudspeaker array. Loudspeakers were selected as the nearest to equal spacing

JASA
https://doi.org/10.1121/10.0030462

across the array. Figure 1 illustrates how the weights for individual loudspeakers sampled values from the underlying function in our experiments using Eq. (2). In certain cases, it is possible for the weighting function to place all weight on a single loudspeaker. In experiment 1, the loudspeaker at 0° was always active and the other active loudspeakers were spaced as equally as possible over the rest of the array. This arrangement avoided the presentation of point sources in which the two sound sources were spatially separated, but in the baseline condition, target and masker were presented only from the speaker at 0°. In a specific, point-source condition, only the loudspeakers at the exact azimuths were activated, and this condition was treated as representing an ambisonic order of infinity:

$$f(\theta, m) = \sin\left[\frac{\sin\left(\frac{2m+1}{2}\theta\right)}{n\sin\left(\frac{\theta}{2}\right)}\right], \tag{1}$$

$$f(\theta, m) = \sin\left[\frac{\sin\left(\frac{2m+1}{2}\theta\right)}{(2m+1)\sin\left(\frac{\theta}{2}\right)}\right]. \tag{2}$$

Target and interfering stimuli were presented in three spatial configurations at ±15°, ±30°, and ±45°. The target was always located to the listener's right. Symmetric configurations were used because they produce the largest SRM. The ambisonic rendering used orders of 1, 2, 4, 8, and infinite (point sources). The baseline condition for calculation of the SRM had target and masker at 0°. With 3 spatialized conditions × 5 ambisonic orders and 1 baseline condition, there were 16 conditions in all.

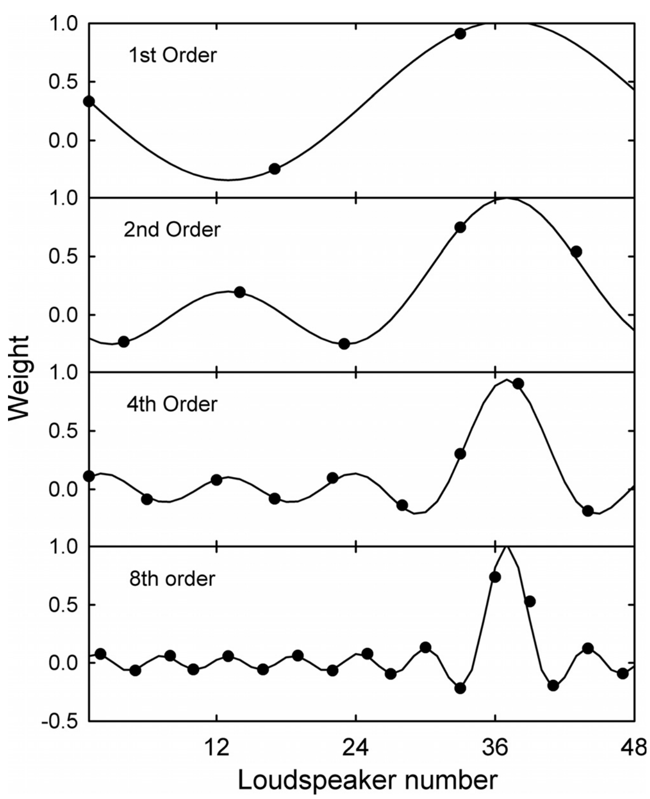

FIG. 1. Illustration of ambisonic panning for our 48-loudspeaker ring. The lines are weighting functions for different orders of ambisonic panning for a source at 30°. Loudspeaker 32 is at 0°. The filled circles are the weightings of individual loudspeakers, which are sampled from the function.

C. Participants

The participants were 16 undergraduates at Cardiff University with English as first language and no known hearing impairment.

D. Procedure

The participants attended a single, 45-min session. They were seated in the center of the loudspeaker array with the experimenter located behind them and outside the loudspeaker ring with a computer monitor and keyboard. The participant's head position was verified by running a cord between the loudspeakers at ±90° over the top of the head and visually aligning it with the interaural axis. Each participant was instructed to face the front and listen for speech in each stimulus, reporting any words that they heard. The experimenter scored the listener's report on the keyboard with reference to the correct transcript displayed on the monitor.

Each measurement began with the speech presented at a signal-to-noise ratio (SNR) of –30 dB such that no words would initially be heard. After each trial, the listener reported verbally to the experimenter what they had heard. The stimulus was repeated with the SNR increased by 4 dB until the listener was able to correctly report at least two of the five keywords from the first sentence in a list. Once this threshold was crossed, the SNR was adjusted adaptively, and each trial used a new sentence from the list. The SNR was increased by 2 dB if the listener heard three or more words correctly and decreased by 2 dB if this criterion was not met. Trials continued until all ten sentences had been presented and the speech reception threshold (SRT) was the average of the last seven SNRs presented plus the next SNR that would have been used given the result with the tenth sentence.

There was one practice measurement using target and interferer at 0°, employing first-order ambisonics. The result of the practice was discarded. The 16 conditions were then presented in a random order using 16 IEEE sentence lists. The order of the conditions was rotated with each successive participant while the sentence-list order remained fixed. This procedure counterbalanced the effects of learning and fatigue on the different conditions at the same time as the intrinsic intelligibility variations of the different sentence lists.

E. Results

The SRTs from each condition were subtracted from the SRT in the baseline condition for that participant to yield measures of SRM. The mean resulting SRMs, averaged across participants, are displayed in Fig. 2 and were analysed with a 3×5 repeated-measures analysis of variance. Figure 2 shows a progressive increase in SRM as the azimuthal separation of the target and interferer increases, resulting in a main effect of spatial configuration [$F(2,30) = 70$, $p < 0.001$]. There is also a progressive increase in SRM with increasing ambisonic order [$F(4,60) = 21$, $p < 0.001$]. There was no significant interaction between these two effects [$F(8,120) = 1.09$]. The effect of ambisonic order was interrogated with *post hoc* Holm-Bonferroni-corrected *t*-tests ($\alpha = 0.05$). These showed that each level of ambisonic order differed from every other with the exceptions of 2 vs 4, 4 vs 8, and 8 vs infinity.

F. Discussion

Because fourth order differs significantly from infinity, these results indicate that SRM increases with ambisonic order beyond an order of four. Judging from Fig. 2, eighth-order ambisonics appear to be almost indistinguishable from infinite order, suggesting that asymptote with real point-source presentation (infinite order) occurs somewhere toward or around eighth order. This outcome seems broadly consistent with the results of Ahrens *et al.* (2020), who found progressive improvement in SRM up to fifth order but not beyond. The effect of ambisonic order is evident for each of the spatial separations with no significant interaction between the two variables, indicating that a narrow spatial separation is not necessary to demonstrate benefits of higher-order ambisonics.

III. EXPERIMENT 2

Experiment 1 showed that the ambisonic panning technique can display progressive benefits of increasing ambisonic order up to at least fourth order. However, similar to Ahrens *et al.* (2020), the experiment only examined the process of rendering ambisonic signals on a circular loudspeaker array. Experiment 2 extended the evaluation of higher-order ambisonics to include the encoding process as well as examine the influence of room reverberation.

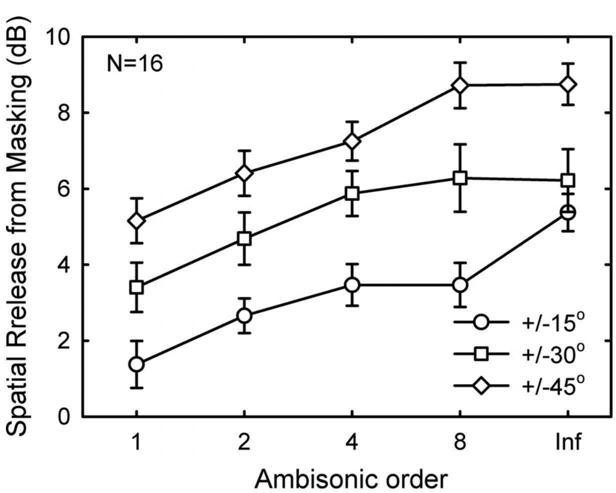

FIG. 2. Data from experiment 1. SRM is depicted as a function of ambisonic order for three angular separations of speech and interfering noise. Ambisonic order was controlled using ambisonic panning. Error bars are one standard error of the mean.

A. Measurement and processing of room impulse responses

A 64-channel ambisonic microphone (em64, mh acoustics, Summit, NJ) was used to record room impulse responses. One room was the sound-deadened laboratory described above ($T60 \approx 60$ ms) and the other room was a 30-m^2 office with no suspended ceiling ($T60 = 350$ ms). For this purpose, a Mission M30i loudspeaker (Huntingdon, UK) was used to present tone sweeps generated by a Dell laptop (Round Rock, TX) with a Realtek® soundcard (Hsinchu, Taiwan).

The ambisonic impulse responses were recorded from two directions, differing by 90° in each of the two recording environments. The impulse responses were measured using the tone-sweep method described by Farina (2007). The 10-s tone sweeps were presented from a distance of 1.2 m (matching the radius of the loudspeaker ring) and ranged from 65 Hz to 20 kHz. To match the conditions of experiment 1 as closely as possible, the 64 raw microphone outputs were recorded by Eigenstudio 3 software from mh acoustics (Summit, NJ) and subsequently encoded into the first-, second-, fourth-, and sixth-order ambisonics using the same software. Once encoded into *B*-format, the recordings were downsampled from 48 to 44.1 kHz. They were then spatially rotated so that they would be aligned to ±45° on the loudspeaker array. Decoding into impulse responses for each loudspeaker was performed using a sampling decoder. The rotation and decoding were performed using open-source MATLAB software (Politis, 2016).

B. Stimuli

IEEE sentences from the same talker were used as target stimuli while interfering noise was generated in the same manner as in experiment 1. The speech and noise signals were convolved with the appropriate impulse responses for each loudspeaker to render them onto the loudspeaker array. In experiment 2, the loudspeaker at 127.5° was always active, and the other active loudspeakers were spaced as equally as possible over the rest of the array.

Unlike experiment 1, the speech was not always located on the same side. Instead, separate measurements were made with speech on the left or right with the noise in the other location. These two measurements were then averaged. This refinement was implemented to account for any effect of asymmetry in the loudspeaker array or that caused by the residual acoustic reflections in the listening room.

C. Participants

The participants were 22 undergraduate students from Cardiff University with English as first language and no known hearing impairment.

D. Procedure

The procedure was similar to that of experiment 1. With 4 ambisonic orders (1st, 2nd, 4rd, and 6th), 2 listening environments, and 2 mirror-image spatial configurations for each condition, there were 16 SRTs for spatialized speech and noise. With the addition of baseline, non-spatialized SRTs for each environment (both sources at 0°) and four other conditions that are not reported here,[1] there were 22 SRTs in the experiment. In a refinement to the method of experiment 1, the participant's head was precisely positioned in the center of the ring using ceiling-mounted, laser crosshairs that were aligned to the loudspeakers in the cardinal directions.

After the practice run, 22 SRTs were measured in a random order during each 75-min session. As in experiment 1, the order of the conditions was rotated to counterbalance order and material effects. An initial SNR of −21 dB was used for each measurement.

E. Results

As in experiment 1, SRTs for each condition were subtracted from the corresponding baseline SRT to give measures of SRM. The mirror-image configurations were then averaged, leading to a single SRM for each of the five ambisonic orders and two environments. The mean SRMs for 22 listeners are plotted in Fig. 3. As in experiment 1, the SRM increases progressively with ambisonic order. The results from Dagan *et al.* (2019), using the same spatial configuration and noise type, are superimposed for comparison.

The results were analysed with a 2×4 analysis of variance for SRM, covering the two listening environments and four levels of ambisonic order. There was a significant main effect of ambisonic order [$F(3,63) = 285$, $p < 0.001$] and a significant main effect of listening environment [$F(1,21) = 6.5$, $p < 0.05$] but no interaction between the two [$F(3,63) = 0.9$]. Comparisons of different levels of ambisonic order using Holm-Bonferroni correction revealed that all levels differed from each other ($p < 0.001$, in each case) except for orders 4 and 6.

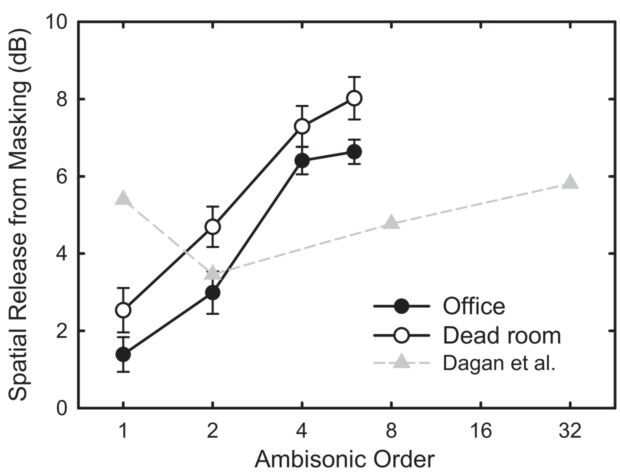

FIG. 3. Data from experiment 2. SRM as a function of ambisonic order for two different rooms using an ambisonic microphone. Error bars are one standard error of the mean. Data from Dagan *et al.* (2019) were scanned from the 50% point of the fitted psychometric functions in their Fig. 3.

F. Discussion

Experiment 2 tested whether SRM increases with ambisonic order when an end-to-end system is used. The system included the microphone recording and encoding process and the decoding of the resulting *B*-format signal onto the loudspeakers. The results agree with those of experiment 1 and Ahrens *et al.* (2020) in showing a progressive increase in SRM with ambisonic order. The results indicate that most of the improvement in SRM with ambisonic order is achieved with fourth order, which yielded 7.0 dB in the acoustically dry laboratory, whereas sixth order produced 7.8 dB. To check whether these values are as expected, we used binaural room impulse responses (BRIRs) recorded over KEMAR (Burkhard and Sachs, 1975) in the same listening space to generate a theoretical prediction from the model by Jelfs *et al.* (2011) for this configuration. The BRIRs were recorded using a 10-s tone sweep (Farina, 2007). This theoretical prediction was 10.2 dB, which is a little higher than that observed with sixth-order ambisonics.

IV. GENERAL DISCUSSION

Both experiments showed that SRM improves progressively with higher-ambisonic orders. Most of the benefit was achieved with an ambisonic order of four, but there was evidence of continued improvement through sixth and eighth order. Moreover, the first experiment found no evidence that the effect was moderated by the size of the spatial separation, and the second experiment found no evidence of any interaction with room acoustics. The benefit of higher-order ambisonics is, thus, not limited to high-fidelity listening environments or narrowly separated sound sources.

A. Comparison of the two experiments

The two experiments used quite different methods for generating an ambisonic sound field. Experiment 1 used an ambisonic panning method that simply applies a scale factor to a source signal at each loudspeaker. Experiment 2 used a high-order ambisonic microphone. In the latter case, the *B*-format impulse responses were then decoded onto the loudspeaker configuration and convolved with the desired source sounds. Although these methods seem quite distinct, Fig. 4 illustrates the fact that once the impulse responses from a second-order ambisonic encoding by the em64 (mh acoustics, Summit, NJ) were decoded onto the loudspeaker configuration, the five different waveforms were essentially scaled versions of each other. Thus, both methods are delivering scaled versions of the source waveform to the different loudspeakers.

As both experiments investigated the case of having sound sources at ±45°, we can directly compare the SRM values across the two experiments for that case. The laboratory condition from experiment 2 involved impulse

https://doi.org/10.1121/10.0030462

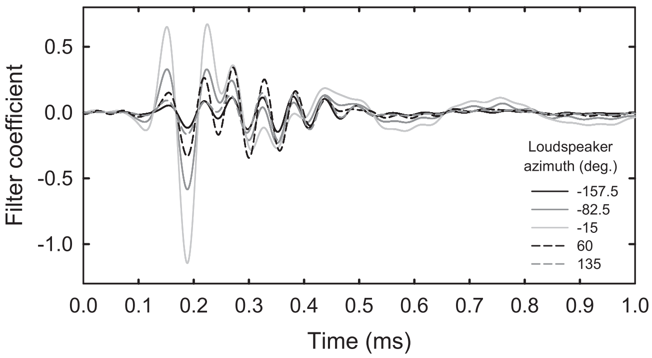

FIG. 4. Example set of second-order impulse responses for a source at –45° in the sound-treated laboratory and decoded onto five loudspeaker positions. For clarity, the impulse responses have been ten times upsampled.

responses with very little reverberation, hence, its results should correspond with those from the ambisonic panning method in experiment 1. Table I shows that the SRM values are consistently lower in the microphone-based conditions. Some of this deficit could be caused by the residual reverberation in the sound-deadened laboratory, which would have been encoded during the process of recording the impulse responses and present during playback over the loudspeaker ring.

B. Comparison with previous studies

The results of the two experiments seem consistent with the findings of Ahrens *et al.* (2020) that SRM is sensitive to ambisonic order. It also extends the observation of Ahrens *et al.* (2019) that seventh-order ambisonics was sufficient to reproduce the full effect of SRM in a reverberant space by showing the asymptotic approach to this case with increasing ambisonic order. The studies by Ahrens *et al.* (2020) and Ahrens *et al.* (2019) used speech interferers and, as such, are not directly comparable with the present results, but the trends are nonetheless similar. In contrast, the use of speech-shaped noise as an interferer in our experiments should make them directly comparable with the results of Dagan *et al.* (2019), even though other aspects of the implementation of ambisonics, such as their use of headphones, may have differed. Despite this, the results appear to contrast quite strongly.

Dagan *et al.* (2019) concluded that first-order ambisonics were sufficient to yield the full effect of SRM. Our results from two experiments with different underlying methodologies led us to question this conclusion. As noted above, the maximum size of SRM that they observed falls substantially short of the theoretical expectation from the

TABLE I. SRM (dB) measured in each of the experiments for three levels of ambisonic order and the ±45° configuration.

Ambisonic order	Experiment 1	Experiment 2 (Dead room)
1	5.16	2.53
2	6.41	4.69
4	7.25	7.30

model by Jelfs *et al.* (2011). We can now add that it also falls short of the maximum levels of SRM that we observed in the present experiments using the same spatial configuration and same type of interferer. This contrast is highlighted in Fig. 3, where we have superimposed SRM values derived from the paper by Dagan *et al.* (2019). Whereas our experiment seems similar to that by Dagan *et al.*, there remain a number of differences that might explain the outcome. These include the possible effect of headphone presentation vs presentation with a free head in an ambisonic sound field and the algorithm for selecting the directions of the active loudspeakers. As noted in the Introduction, it is possible to select those angles such that the sound field contains only point sources in a given spatial configuration. The existence of this extreme case may be indicative of smaller effects that may occur as the loudspeaker angles are rotated around the listener.

C. Modelling the experiments

The two experiments were modeled using the model of SRM by Jelfs *et al.* (2011), which is based on processing BRIRs. In previous work, the model has proved successful at recreating the pattern of results for SRM experiments with steady-state interfering noises in artificial and real listening spaces, but these experiments have always employed a single target source and some number of independent interfering sound sources. To predict experiments with multiple independent interferers, the BRIRs for each interferer were concatenated rather than added together because summing impulse responses will create phase interactions that would not exist for independent sound sources. In the case of ambisonic presentation, however, target and interferer(s) are rendered using multiple loudspeakers that carry the same signal with different scale factors. In this case, therefore, directly summing the BRIRs for a given source should be appropriate as it will capture the phase interactions between the coherent signals from each loudspeaker. The modelling presented here is the first attempt to use the model in this way for coherent sound sources, but the situation is very similar to the established and validated methods of modelling sound reflections, which are also coherent with the source and summed together to form the BRIR (Jelfs *et al.*, 2011).

BRIRs were measured using 10-s tone sweeps between each of the 48 loudspeakers and a KEMAR placed at the center of the array. The KEMAR was precisely positioned using the laser crosshairs. As the waveforms were to be summed, it was important to conserve the time of flight in each recording, which is not usually essential in BRIR measurement. This was performed by using the same audio interface (Motu 24I/O Core, Cambridge, MA) for synchronous input and output.

In experiment 1, sound sources were directly presented within the sound-treated laboratory. To model this case, the laboratory BRIRs from each of the active loudspeakers were scaled and summed separately for each of the two sound sources, target and interferer, before delivering them to the model. Each ambisonic order, thus, involved summing different numbers of BRIRs. Figure 5 shows the pattern of predicted thresholds (lines) combined with the mean SRM measurements from the experiment (symbols). The correlation between data and prediction is 0.91, which is at the low end of such correlations in previous work (Culling *et al.*, 2013). However, the model captured the general trend toward higher SRM with increased ambisonic order and increased spatial separation. The former effect appears to be partially non-monotonic in the model with a dip in the prediction for the $\pm 15°$ and $\pm 30°$ configurations at an ambisonic order of eight. The model is, therefore, in broad agreement with the experimental result that SRM increases with spatial separation and ambisonic order, but it is unclear why the predicted effect of ambisonic order is sometimes non-monotonic, which is largely produced by changes in the predicted effect of better-ear listening. The model also tends to predict somewhat larger values of SRM than were measured, by an average of 1.6 dB.

In experiment 2, sound sources were initially presented in a reverberant room or within the sound-treated laboratory to the ambisonic microphone. These recordings were then rendered in the laboratory such that, effectively, the sounds all passed through two rooms before reaching the participant. To model this case, the spherical-harmonic impulse responses for the two locations at $\pm 45°$ were rendered onto loudspeaker channels just as the *B*-format signals had been in the experiment (including the rotations). They were then convolved with the corresponding loudspeaker BRIRs and summed at each ear to produce binaural impulse responses for the entire signal path. These impulses responses for target and masker were presented to the model. Because mirror-image locations had been used in experiment 2, the predictions for these two spatial configurations were averaged to produce Fig. 6. For experiment 2, the model correctly predicts that SRM increases monotonically with

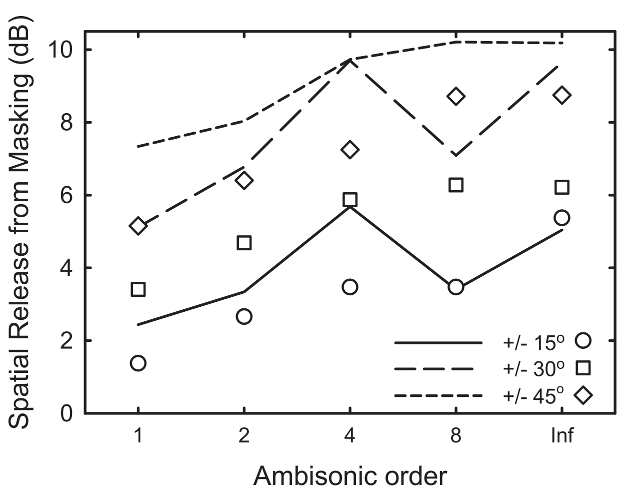

FIG. 5. Comparison of model-predicted and observed SRMs for experiment 1 using the model by Jelfs *et al.* (2011). Lines are model predictions, and symbols are mean SRM values, reproduced from Fig. 2.

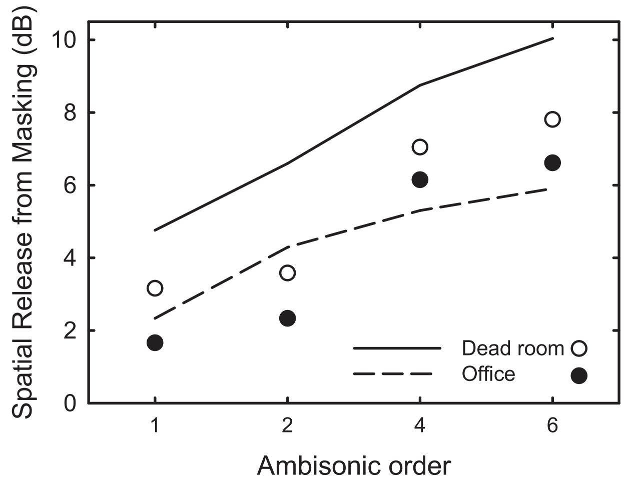

FIG. 6. Comparison of model-predicted and observed SRMs in experiment 2 using the model by Jelfs *et al.* (2011). Lines are model predictions and symbols are mean SRM values, reproduced from Fig. 3.

ambisonic order and SRM is generally greater for the sound-deadened-laboratory condition compared to the reverberant-office condition. As with experiment 1, the correlation between prediction and data was moderate (0.86), and the predicted values tended to be higher than the observed values by an average of 1.0 dB.

D. Conclusions

Ambisonic order has a strong effect on SRM for speech presented against noise. The effect can be observed in sound-deadened and reverberant environments, up to at least fourth order and for a range of different spatial separations. The model of SRM by Jelfs *et al.* (2011) captures the main trends in these results.

ACKNOWLEDGMENTS

Work supported by U.K. Engineering and Physical Sciences Research Council. For the purpose of Open Access, the author has applied a CC BY public copyright license to any author accepted manuscript (AAM) version arising from this submission.

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Ethics Approval

The procedures for both experiments in this work were approved by the School of Psychology Ethics Committee at Cardiff University in compliance with the Declaration of Helsinki.

DATA AVAILABILITY

The data that support the findings of this study are available from the corresponding author upon reasonable request.

[1]Four other conditions were originally intended to parallel the "infinite-order" conditions from experiment 1, but it was later realised that using two-point sources was not consistent with reproducing reverberation.

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.
Ahrens, A., Marschall, M., and Dau, T. (**2020**). "The effect of spatial energy spread on sound image size and speech intelligibility," J. Acoust. Soc. Am. **147**, 1368–1378.
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.
Brungart, D. S. (**2001**). "Informational and energetic masking effects in the perception of two simultaneous talkers," J. Acoust. Soc. Am. **109**, 1101–1109.
Burkhard, M. D., and Sachs, R. M. (**1975**). "Anthropometric manikin for acoustic research," J. Acoust. Soc. Am. **58**, 214–222.
Culling, J. F., Lavandier, M., and Jelfs, S. (**2013**). "Predicting binaural speech intelligibility in architectural acoustics," in *The Technology of Binaural Listening*, edited by J. Blauert (Springer, Heidelberg).
Dagan, G., Shabtai, N. R., and Rafaely, B. (**2019**). "Spatial release from masking for binaural reproduction of speech in noise with varying spherical harmonics order," Appl. Acoust. **156**, 258–261.
Farina, A. (**2007**). "Advancements in impulse response measurements by sine sweeps," in *Proceedings of 122nd AES Convention*, May 5-8th May, Vienna, Austria (Audio Engineering Society, New York).
Favrot, S., and Buchholz, J. M. (**2009**). "LoRA: A loudspeaker-based room auralization system," Acust. united Act. Acust. **96**, 364–375.
Gardner, W. G., and Martin, K. D. (**1995**). "HRTF measurements of a KEMAR," J. Acoust. Soc. Am. **97**, 3907–3908.
Gerzon, M. A. (**1973**). "Periphony: With-height sound reproduction," J. Audio Eng. Soc. **21**, 2–10, available at https://aes2.org/publications/elibrary-page/?id=2012.
Huisman, T., Ahrens, A., and MacDonald, E. (**2021**). "Ambisonics sound source localization with varying amount of visual information in virtual reality," Front. Virtual Real. **2**, 7223221.
Jelfs, S., Culling, J. F., and Lavandier, M. (**2011**). "Revision and validation of a binaural model for speech intelligibility in noise," Hear. Res. **275**, 96–104.
McCormack, L., Pulkki, V., Politis, A., Scheuregger, O., and Marchall, M. (**2020**). "Higher-order spatial impulse response rendering: Investigating the perceived effects of spherical order, dedicated diffuse rendering, and frequency resolution," J. Audio Eng. Soc. **68**, 338–354.
Moore, B. C. J., and Glasberg, B. R. (**1983**). "Suggested formulae for calculating auditory-filter bandwidths and excitation patterns," J. Acoust. Soc. Am. **74**, 750–753.
Neukom, M. (**2007**). "Ambisonic panning," in *123rd Convention of Audio Engineering Society* (Audio Engineering Society, New York), paper 7297.
Noisternig, M., Musil, T., Sontacchi, A., and Holdrich, R. (**2003**). "3D binaural sound reproduction using a virtual ambisonic approach," in *IEEE International Symposium on Virtual Environments, Human-Computer Interfaces and Measurement Systems, 2003, VECIMS '03*, July 27–29, Lugano, Switzerland (IEEE, New York), pp. 174–178.
Politis, A. (**2016**). "Microphone array processing for parametric spatial audio techniques," thesis, Aalto University, Espoo, Finland.
Rothauser, E. H., Chapman, W. D., Guttman, N., Nordby, K. S., Silbiger, H. R., Urbanek, G. E., and Weinstock, M. (**1969**). "I.E.E.E. recommended practice for speech quality measurements," IEEE Trans. Aud. Electroacoust. **17**, 227–246.
Stitt, P., Bertet, S., and van Walstijn, M. (**2014**). "Off-centre localisation performance of ambisonics and HOA for large and small loudspeaker array radii," Acta Acust. Acust. **100**, 937–944.
Wagener, K., Josvassen, J. L., and Ardenkjaer, R. (**2009**). "Design, optimization and evaluation of a Danish sentence test in noise," Int. J. Audiol. **42**, 10–17.
Zotter, F., and Frank, M. (**2019**). *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality* (Springer, Cham, Switzerland).