Kim, Taeho; Pulkki, Ville

Median Plane Localization of Short Noise Bursts in the Presence of Horizontally Spread White Noise Masker

Published in:
AES: Journal of the Audio Engineering Society

DOI:
10.17743/jaes.2022.0172

Published: 01/11/2024

Document Version
Peer-reviewed accepted author manuscript, also known as Final accepted manuscript or Post-print

Please cite the original version:
Kim, T., & Pulkki, V. (2024). Median Plane Localization of Short Noise Bursts in the Presence of Horizontally Spread White Noise Masker. *AES: Journal of the Audio Engineering Society*, *72*(11), 776-796. https://doi.org/10.17743/jaes.2022.0172

Median plane localisation of short noise bursts in the presence of horizontally-spread white noise masker

Taeho Kim & Ville Pulkki
Aalto University
Aalto Acoustics Lab.
Department of Information and Communications Engineering

Abstract

Recent spatial audio techniques involve separating multichannel signals into direct and background parts. However, determining parameters for localising short sound sources in background sounds remains challenging due to the limited knowledge of spatial hearing resolution. This paper investigates the localisation performance when short bursts in the median plane are presented with spectrally similar, horizontally spread broadband noise. Listening tests examined target stimuli in different median plane locations with and without masker noises, using elevation gain, bias, and error rate to evaluate localisation performance. The target stimuli comprised aperiodically repeated multiple-burst stimuli with different burst rates and levels, and single-burst stimuli with varied duration and levels. The results showed that the burst rate of multiple-burst stimuli had a weak systematic effect on all the criteria for localisation performance, regardless of noise. However, extending the duration of single-burst stimuli increased the elevation gain, and the added noise further improved the localisation performance. The masker also improved localisation performance when the sound level increased, while unmasked stimuli had the opposite effect. The optimal conditions for improving localisation performance with background noise found in this study were a signal-to-noise ratio of $\geq 18\,\text{dB}$.

1 INTRODUCTION

The separation of recorded multichannel signals into contributions of direct and reverberated or ambient parts has been of interest lately in the context of time-frequency-domain parametric spatial audio techniques [33]. In such processing, it is of vital importance to choose the parameters for the separation process following the human resolution of spatial hearing. Although several parameters can be deduced from existing knowledge, the information for resolution and accuracy of localisation of short sound sources in the presence of spatially extended background sounds is lacking. Thus, studying the aspects of human auditory localisation of short sound in background noise is necessary to determine the appropriate parameters for the separation techniques in time-frequency-domain parametric spatial audio. This paper explores the optimal stimulus conditions of short noise bursts to achieve precise localisation in the median plane in the presence of background sound.

The localisation of short sound sources in the presence of background noise can be studied by examining a few specific aspects of human perception, such as auditory localisation in background noise and localisation of brief sounds. Auditory localisation of a sound source, while masked by background noise, has been widely studied [e.g., 3, 6, 7, 12, 14, 15, 20, 24, 26, 35]. However, most of these studies have focused on sound sources and noises located in the horizontal plane (an imaginary auditory plane that divides the head horizontally into the upper and lower parts), leaving the localisation of sound sources in the median plane (the auditory mid-line between the ears that vertically divides the head) largely unexplored. Localisation of short sounds has been investigated in several studies [10, 17, 18, 27, 38], reporting that humans can localise short sounds (e.g., duration < 100 ms) in the horizontal plane as accurately as longer sounds. However, they found that localisation performance in the median plane deteriorates as sound duration shortens or intensity increases beyond 80 dB SPL [17]. Nevertheless, the impact of background noise on the localisation of short sounds in the median plane has been rarely studied.

Therefore, this paper focuses on the "median plane localisation" of short sound sources and investigates the effect of added background noise on it. Thus, the perceptual mechanisms are studied using exploratory subjective tests. The experiment examines whether a horizontally spread background noise, spectrally similar to the target noise bursts, could improve the median plane localisation performance of a sound burst or deteriorate it, depending on their duration and signal-to-noise ratio (SNR). In addition to the single-burst stimuli, the localisation of multiple noise bursts of different average burst rates was also examined with various conditions. The average burst rates were adjusted to allow the listeners to detect every single burst, and the bursts were temporally distributed to be aperiodic to mimic the random appearance of multiple glimpses of sound. This experiment setup focused on finding whether multiple occurrences of detectable glimpses can improve localisation performance than a short single burst.

Since this study aims to find how masking noise affects the ability to localise short noise bursts in the median plane under different conditions, the experiment first validates earlier studies using specific criteria to assess localisation performance, including elevation gain, elevation bias, and response error rates. Then, the effect of spatially extended background noise is examined on localisation performance, which is presented with short sound stimuli of various conditions, such as:

1. Effects of noise on multiple-burst stimuli with different burst rates.
2. Effects of noise on single-burst stimuli with different duration.
3. Effects of noise on multiple-burst stimuli with different sound levels (SNRs).
4. Effects of noise on single-burst stimuli with different sound levels (SNRs).

Consequently, the findings from our perceptual studies can be used to determine optimal stimulus conditions for time-frequency-domain parametric spatial audio techniques that can separate multichannel signals into direct and ambient parts based on spatial hearing resolution. In addition, since most natural sound scenes include background and ambient sounds, the obtained parameters can help create spatial audio content that offers listeners more precise sound localisation and can also guide the development of spatial audio applications that deliver an advanced localisation experience.

2 BACKGROUND

Human sound localisation relies on implicit acoustic cues such as binaural and monaural cues. Binaural cues arise from interaural disparities in the sound level and arrival time between each ear, which are determined via the *interaural level difference* (ILD) and the *interaural time difference* (ITD), respectively [29, 34]. They allow the localisation of the azimuth angle in the horizontal plane. Monaural cues indicate direction-dependent spectral filtering mainly caused by the reflections and diffraction of soundwaves due to the head, torso, and pinnae. These *spectral* cues are fundamental to the localisation in the vertical directions and front-back differentiation especially in the median plane, where the interaural disparities (binaural cues) are at their minimum [4, 29, 34]. Combining the binaural and monaural (spectral) cues contributes to localisation in three dimensions [34].

The effect of sound masking on auditory localisation has been examined in relation to the SNR or the locations of a target signal and a masker noise. According to multiple studies, the accuracy of localisation often decreases as the SNR decreases [6, 14, 26, 41]. Furthermore, Zwiers et al. (2001) [41] and Ege et al. (2018) [12] reported that the response gain (the slope of the fitted line describing the relation between the stimuli and response locations) and the correlation between the stimuli and responses decreased as the SNR decreased. Besides, Lorenzi et al. (1999) [26] reported that the localisation accuracy tended to be unaffected by noise until the SNR was 0–6 dB.

However, conflicting results about the effect of signal and masker locations have also been reported. In the horizontal directions, the perceived direction of a signal was either biased toward the masker (*pulling effect*) [e.g., 15] or away from the masker (*pushing effect*) [e.g., 3, 7]. Brungart and Simpson (2009) [6] reported that the perceived direction depended on the relative difficulty of localisation. An easier localisation condition (e.g., high SNR) tended to result in pushing effects, whereas a more difficult one (e.g., low SNR) resulted in pulling effects. On the other hand, the effect of a masker noise on vertical localisation was reported to be stronger than horizontal localisation [6, 41] but barely systematic [3, 24].

It is well known that the binaural cues are robust and precise in horizontal directions. Additionally, median plane localisation is also known to be moderately accurate, provided that a broadband sound source with sufficient duration and sound level is presented to listeners with normal hearing [e.g., 1, 8, 23, 28, 39]. However, the localisation of sound sources under other conditions (e.g., duration is less than 100 ms) has not been widely investigated. Several studies examined the localisation of brief sound stimuli [10, 17, 18, 27, 38]. They mainly focused on vertical directions since sounds of short duration (< 100 ms) were found to be as accurately localised as longer sounds in the horizontal directions [e.g., 18, 27]. Hofman and Van Opstal (1998) [18] reported that a 3-ms sound stimulus was localised as accurately as a 500-ms stimulus. Additionally, Macpherson and Middlebrooks (2000) [27] found little effect of the duration or level of sound stimuli on horizontal localisation.

However, in vertical directions, the localisation accuracy was reported to vary depending on the sound stimuli's duration or level. The localisation performances varying with stimulus duration can be associated with the perceived resolution of spectral cues. Since spectral cues are the primary factor for the localisation in the median plane, the auditory system estimates the direction of a sound source by extracting average spectral shapes of direction-dependent filters (e.g., head-related transfer functions [30]) from the received spectra at the eardrums. Accordingly, the "estimation" process might require a minimum source duration to secure sufficient spectral detail for the localisation. Since the spectral resolution necessary for precise localisation has been investigated in only a few studies [16, 22], the hypothesis regarding the minimum source duration motivates studying median plane localisation of short-duration stimuli.

Hofman and Van Opstal (1998) [18] studied the effects of various spectro-temporal factors on both horizontal and vertical sound localisation in the frontal hemifield. First, they examined the perceived elevation with short white Gaussian noise bursts ranging from 3 ms to 80 ms. The "elevation gain", representing the fitted line's slope on the relation between the stimulus direction and perceived elevation, was applied to systematically analyse the results. Whereas the 80 ms stimulus led to an elevation response gain close to the ideal value (unity), the gain declined with decreasing stimulus duration. From this result, Hofman and Van Opstal suggested that at least an 80-ms duration is necessary for the auditory system to "integrate" a broadband sound input to obtain a stable elevation estimation.

However, Macpherson and Middlebrooks (2000) [27] claimed that the deterioration of elevation gain from short noise bursts is more related to the *negative level effect* from high-level stimuli than the deficiency of integration quality in the auditory system. The *negative level effect*, reported by Hartmann and Rakerd (1993) [17], refers to the phenomenon that, when localising a click train in the median plane, localisation error starts to increase with stimuli level of about 80 dB SPL and above. They suggested that this phenomenon is due to the saturation of the peripheral excitation in the cochlea for high-level stimuli. In Hartmann and Rakerd's research, the error rate of overhead stimuli was higher than the front stimuli for high-level clicks, which indicates a decrease in elevation gain in the frontal hemifield.

Macpherson and Middlebrooks further examined this phenomenon by comparing the changes in the elevation gain for high and low SLs (sensation level: the level above the auditory threshold) with 3-ms and 100-ms noise bursts. Although the elevation gain from the short noise burst (3 ms) decreased as the stimuli level increased from around 25 dB to 60 dB SL, the 100-ms burst showed no significant changes in elevation gain at different levels. They explained this observed difference due to the auditory system adapting to high-level stimuli when they are provided with sound of sufficient duration. The latter portion of sufficiently long stimuli is thought to spare the cochlea saturation caused by high-level sounds. As a result, the earlier portion of the stimulus may activate an adaptive mechanism.

Macpherson and Middlebrooks also investigated this hypothesis by co-presenting a 1000-ms diffuse noise with a short noise burst (3 ms) centred within the noise that would enable the high-level stimulus adaptation in the auditory system prior to the onset of the noise burst. As a result, slight rises in the elevation gain were observed when the background noise level increased, validating the adaptation hypothesis. However, this effect appeared in limited conditions. For example, the increase in the elevation gain was only apparent with a high-level stimulus (55 dB SL) above the estimated saturation threshold (40 dB SL). Moreover, the gain only increased at 10 to 20 dB SNRs between the stimulus and the diffuse noise. All other conditions showed decreases in the elevation gain as the SNR decreased.

Vliegen and Van Opstal (2004) [38] examined the median plane localisation of short sounds with differences in both duration and level. They found the duration-gain correlation found by Hofman and Van Opstal (1998) [18], where elevation gains increase with increasing duration of short stimuli, appeared at every stimulus level they tested. Furthermore, the negative level effect above 55–65 dB SPL appeared in every stimulus duration up to 100 ms. In addition, low-level stimuli below 55–65 dB SPL provided increasing elevation gain with the level increase. These findings partly disagree with Macpherson and Middlebrooks (2000) [27], where low-level stimuli and long sound bursts barely showed variability in the elevation gain. However, Vliegen and Van Opstal (2004) [38] suggested that neither neural integration nor peripheral adaptation in the auditory system could determine the elevation gain exclusively. Instead, they might collectively support the median plane localisation.

Burst-train stimuli were also used to investigate the localisation of short-duration sound stimuli [17, 18]. Using the burst-train stimuli, Hofman and Van Opstal (1998) [18] investigated the effect of the silence gap between 3-ms noise bursts on the median plane localisation. The elevation gain was close to unity with 3–10 ms gaps for all the subjects, then decreased monotonically when increasing the silence gap. This result indicated that even bursts with durations of 3 ms were stably localised in the median plane when repeated with a short enough interval. Therefore, they suggested that the directional information could be kept unfaded in the auditory memory within the observed short intervals.

However, the stimuli within these silence gaps might be heard as continuous noise rather than as a series of short-duration stimuli since the burst rates for the burst-train stimuli with 3–10 ms intervals, approximately 80–160 Hz, would result in poor fluctuation strengths due to the temporal masking between the bursts [13]. Considering this phenomenon, it is notable that longer silence gaps (20–80 ms), which might allow the individual bursts to be more detectable, resulted in lower elevation gains than the 3–10-ms gaps, yet higher than the single-burst stimuli.

To summarise, previous studies report that masking noise affects auditory localisation depending on the SNR and the location of target stimuli and the masker. The localisation accuracy declines systematically as the SNR decreases, and the perceived location of the stimulus is either pushed away from or pooled towards the masker location, depending on the perception difficulty. Furthermore, the localisation performance of short-duration sound stimuli in the median plane deteriorates as the stimulus duration shortens and the intensity increases above a certain high level. However, the influence of masking noise on short-duration stimuli has not been studied extensively. As mentioned earlier, Macpherson and Middlebrooks (2000) [27] reported a slight increase in the elevation gain of 3-ms noise burst due to the added background noise. Nonetheless, this high-level stimulus adaptation from a pre-existing noise has not been investigated with different stimulus durations. Thus, this paper examines the accuracy of localising short, vertically located sounds in background noise with varying stimulus durations and SNRs. It also examines the localisation of multiple-burst stimuli comprising detectable single bursts.

3 METHOD

The experiment investigated the localisation characteristics of short noise bursts masked by broadband noise. Median plane localisation was examined by varying the duration, burst rates (per second), SNRs (target-to-masker ratios), and source locations of the target noise bursts. Accordingly, the experiment comprised two listening tests utilising two different types of target stimuli: sets of multiple noise bursts with various burst rates and single noise bursts that differed in duration. To prevent fatigue and adaptation to the experiment design, each subject completed two test sessions on different days. Each session took about 50 to 70 minutes. Before the main test sessions, a preliminary test was conducted to examine the detection thresholds of each target stimulus in order to determine the SNR set-up of the stimuli.

3.1 Preliminary test

An initial listening test was conducted, where the sound pressure level was controlled with an adaptive staircase method [9], to examine the detection thresholds for each target burst masked by broadband noise. Both target bursts and masking noise were created from random white Gaussian noise. Eight subjects, seven males and one female aged between 18 and 48, all reporting normal hearing, participated in the listening test. The participants were staff members or students from Aalto University. The test was conducted in the anechoic chamber where the main tests were conducted, with the same stimuli set-ups as the main tests described in sections 3.2.3, 3.2.2, and 3.3.1.

In the first stimulus, the burst rates were configured to 1, 2, 4, 8, and 16 bursts per second, with each burst lasting 1 ms. In the second stimulus, the burst duration was set to 0.1, 0.3, 1, 3, and 10 ms. The target and masker stimuli were located along the median plane. The target stimulus was positioned at a 30° elevation above the horizontal plane ($\theta = 0°$, $\phi = 30°$), and the masking noise was presented at the horizontal plane ($\theta = 0°$, $\phi = 0°$). For the staircase test procedure, a 2-down 1-up transformed up-down method with a three alternative-forced-choice (3AFC), as suggested by Levitt (1971) [25] and Shelton and Scarrow (1984) [36], was conducted in this listening test.

As shown in Figure 1, the detection thresholds appeared about −9.6 dB relative to the masker level from the target stimuli of 1 ms and lower for the longer stimuli. The overall detection thresholds were relatively higher than previous studies [e.g., 6, 41] because of the shorter stimulus duration and spectral similarity between the target stimuli and the masker. Therefore, the minimum SNR of the stimuli was set to 6 dB (about 10 dB higher than the highest detection threshold level among the participants) for sufficient audibility in the main listening tests.

3.2 Experiment I

The first experiment investigated the auditory localisation of the multiple noise bursts varying with their mean burst rates (per second).

3.2.1 Subjects

Ten subjects, eight males and two females, participated in the listening test. They were students or staff members of Aalto University aged between 18 and 48. All subjects reported normal hearing.

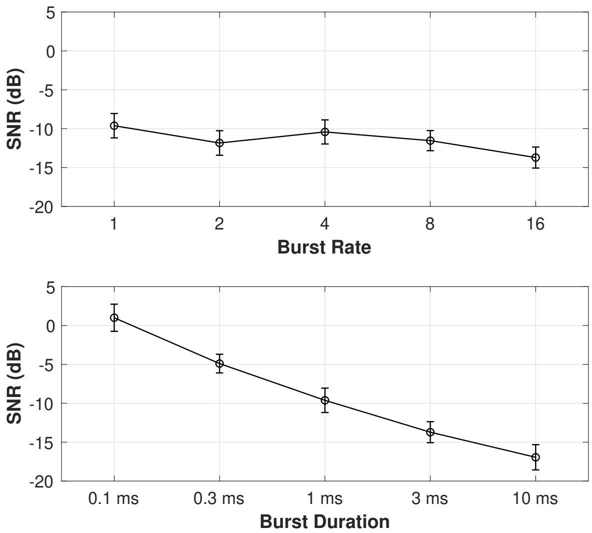

Figure 1: Detection thresholds and standard deviations for two types of target stimuli from the preliminary test. The circles indicate the mean detection thresholds, and the error bars indicate the standard deviations of the response data. The upper plot shows the threshold SNRs and standard deviations of the 1-ms stimuli sets whose mean burst rates range from 1 to 16. The lower plot shows the threshold SNRs and standard deviations of a single stimulus whose duration ranges from 0.1 to 10 ms. The masker level was set to 60 dB SPL, A-weighted.

3.2.2 Apparatus

The localisation test was conducted in an anechoic chamber equipped with forty-six spherically distributed coaxial loudspeakers (Genelec 8331). The listener was positioned in the centre of the chamber. The loudspeakers utilised for this listening test were among those located along the circular auditory mid-line from the listener (the median plane), ranging from $-60°$ to $240°$ polar angles with the interval of $15°$, where the reference ($\phi = 0°$) is positioned in the horizontal plane. The distance from the loudspeakers to the listening position was 2.0 m. Subjects sat on a height-adjustable chair, and their ear level was aligned to the height of the horizontal plane set up for the loudspeakers in the anechoic chamber ($\phi = 0°$). Subjects were instructed to stare at the centre of the coaxial driver of the loudspeaker located at $0°$ azimuth and $0°$ elevation during the listening tests. A headrest was attached to the chair to prevent the subjects' head movements.

3.2.3 Stimuli

White Gaussian noise randomly generated with a 96-kHz sampling rate was utilised for both the masker noise and target bursts. The masker noise source was a 1600 ms

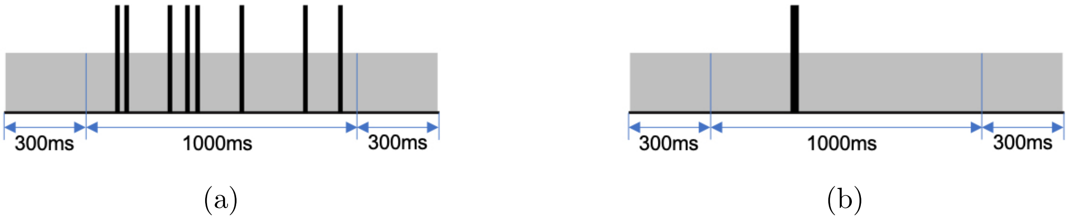

Figure 2: Configuration of target stimuli with the masker. The grey area represents the masker noise, and the short vertical lines represent the target bursts. (a) Stimuli with sets of different burst rates: multiple target bursts, comprising 1-ms single bursts, are randomly distributed within the 300–1300 ms time slot of the masker noise. (b) Stimuli with different burst duration: a single target burst, 1–30 ms in duration, is located in a random temporal position within the 300–1300 ms time slot of the masker noise.

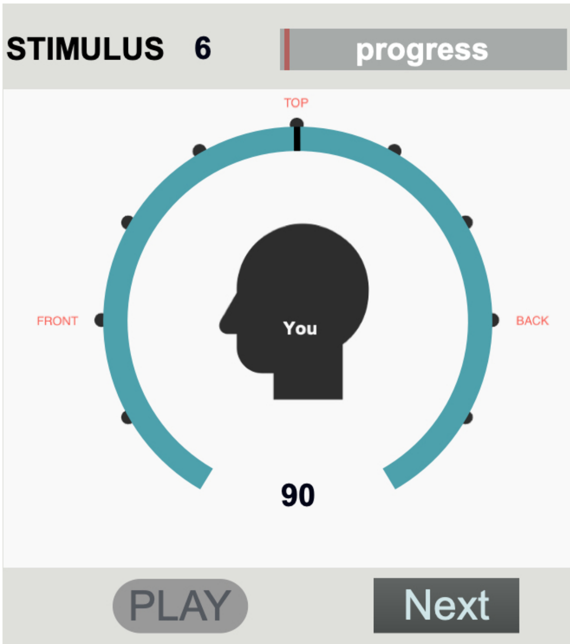

Figure 3: The response interface provided to the subjects. The response scale ranges from −60° to 240°, and the response anchors are displayed from −30° to 210° elevation with the interval of 30°.

long randomly generated white Gaussian noise with 5 ms ramps in their onsets and offsets, and the masker stimuli consisted of two horizontally spread incoherent white noises located at ±30° azimuth to avoid direct masking in the same loudspeaker. The target stimuli were configured as the "sets" of one or more noise bursts, of which the mean burst rates were set to 1, 2, 4, 8, or 16 times per second. The duration of a single target burst was 1 ms, and 0.05-ms onset/offset ramps were applied to each burst. The sound pressure level of the masker pair at the listener position was set to 60 dB SPL, A-weighted, and the SNRs applied to the target stimuli were 6, 10, 14, and 18 dB (66, 70, 74, 78 dB SPL, A-weighted).

As shown in the stimulus configuration shown in Figure 2, the duration of a "set" of bursts was at most 1000 ms, presented between 300 ms and 1300 ms of the masker noise. Thus, when the masker noise was present, a masker-only 300-ms gap was consistently applied to a stimulus's start and end. Furthermore, noise bursts were presented aperiodically with random temporal distribution on the timeline of a burst set. This aperiodic presentation was intended to simulate the random appearance of multiple glimpses.

Significantly, the temporal distribution was configured to recognise every individual burst, preventing any temporal masking between bursts. The distribution of the bursts was carefully controlled to prevent any single burst from being temporally attached or overlapped with other bursts, with the minimum gap between the bursts set to 10 ms, allowing the listeners to hear every 1-ms single target burst. Correspondingly, the authors selected the maximum burst rate of 16 times/sec during pilot tests to secure the perception of temporal "randomness" since burst rates higher than 16 times/sec caused the temporal distribution to be relatively denser than necessary, in which multiple bursts were heard as pseudo-periodic rather than random. Moreover, higher burst rates led the multiple bursts to be perceptually aggregated. In other words, the maximum burst rate was selected to allow all the target stimuli to be heard clearly as an aperiodic repetition of multiple bursts.

3.2.4 Test procedure

The target stimuli were presented to the listener with or without the masker noise during the test. The sound levels of the bursts without the masker noise were identical to the masked bursts. A pair of incoherent masker noise sources were presented from the 30° and −30° azimuth directions in the horizontal plane ($\theta = \pm 30°$, $\phi = 0°$), and the target stimuli were presented from one of seven locations in the median plane: −30°, −15°, 0°, 15°, 30°, 45°, and 60° polar angle. Thus, the target stimulus had various test conditions, including four SNRs, seven target locations, five burst rates, and the presence or absence of the masker noise. Since each stimulus condition was repeated thrice, 840 trials were presented to each subject in pseudo-random order.

A brief training session was carried out before the main test. In the training session, subjects were instructed to use a remote tablet device by presenting a random noise burst at an arbitrary location in the median plane. The response interface displayed a response dial representing the median plane circle. The dial provided a continuous response scale, ranging from −60° to 240° polar angle (see Figure 3).

In the main tests, the task in each trial for the subjects was to press the "PLAY" button in the response interface to play a stimulus and point the dial in the perceived elevation of the target stimuli while staring at the centre (0° elevation) location in the median plane, then proceeded to the next stimulus by pressing the "Next" button. Since the subjects proceeded with the listening test themselves, they could break and rest anytime. Subjects were provided with no prior information about the actual target locations, thus allowing front/back confusion responses (polar angles from 90° to 240° elevation) to be also recorded. They were asked not to move their head during the listening tests to prevent localisation improvement from head movement, which were reported in many studies [e.g. 32, 37, 40].

3.3 Experiment II

The second experiment examined how the burst duration of the target stimuli affects auditory localisation. The experiment involved conducting a listening test in the same venue as the first experiment, with the same subjects and apparatus set-ups.

3.3.1 Stimuli

The masker noise utilised in the second experiment was identical to the first experiment. However, for the target stimuli, single noise bursts of different durations were used in the test. The burst durations applied to the target stimuli were 1 ms, 3 ms, 10 ms, and 30 ms. Each burst was applied with on/offset ramps of 0.05 ms. As shown in Figure 2, a single noise burst was presented randomly within the 300–1300-ms timeline of the masker noise. The sound pressure level of the masker and the SNRs applied to the target stimuli were identical to the first experiment.

3.3.2 Test procedure

Again, the test procedure of the second experiment was identical to the first experiment. The target stimuli were presented to the listener with or without the masker noise during the test. The pair of masker noise stimuli were presented from the 30° and −30° azimuth direction in the horizontal plane, and the target stimuli were presented from one of the seven locations in the median plane (−30°, −15°, 0°, 15°, 30°, 45°, and 60° elevation). The stimuli conditions included seven target locations, four SNRs, four burst durations, and the presence or absence of the masker noise. By repeating each stimulus condition three times, 672 trials were presented to each subject in pseudo-random order.

3.4 Data analyses

Before analysing the results, the front-back reversed responses were distinguished and removed from the main response data. The responses indicating the perceived locations from the rear hemifield (polar angles from 90° to 240° elevation) were considered to be the front-back reversed responses, and the front-back confusion rates were separately analysed (see Figure 9). The response data without front-back reversal were then analysed via 1) the mean values and their confidence intervals, 2) the elevation gains and biases, 3) the signed errors and absolute errors, and 4) the error rates and front-back confusion rate. These values were obtained for each test condition, including burst rate, burst duration, burst level, and target locations. Confidence intervals were derived by applying the 95% confidence level to the means of the responses or errors.

Through linear regression analyses of the collective response data for each test condition, each subject's fitted lines were derived to determine the relationship between the stimulus location and perceived elevation. In the present study, the slope (β) and bias (α) of a fitted line are referred to as the elevation gain and elevation bias, respectively, when the fitted line is represented as $y = \alpha + \beta x$. The elevation gain of 1 represents the ideal fit. These terms have been used as an analytical approach in numerous previous studies [18, 27, 38]. The correlation coefficient (r) and coefficient of determination (R^2) were used to statistically analyse the variations and strength of the elevation gains and biases curves, representing the association between gains/biases and each stimuli condition, including each subject, burst rate, burst duration, level, and the presence of a masker.

Error rates were determined by calculating the percentage of perceived elevations outside the response interval of ±15° or ±30° from the target locations. Front/back confusion rates were also derived from the raw response data.

The authors conducted a statistical analysis to determine whether the masker had a significant effect on responses. They compared the accuracy of responses between unmasked stimuli and stimuli with the masker using ANOVA (Analysis of Variance) and corresponding post hoc tests. Response deviation from the target location (absolute error) was examined for each variable, including burst rate, burst duration, level, and the presence of the masker.

4 RESULTS

4.1 General response trend

Figure 4 presents the means and their confidence intervals obtained from the responses of sound stimuli, both with and without noise, at varying burst rates, burst durations, and burst levels as a function of target locations. Since the masker level

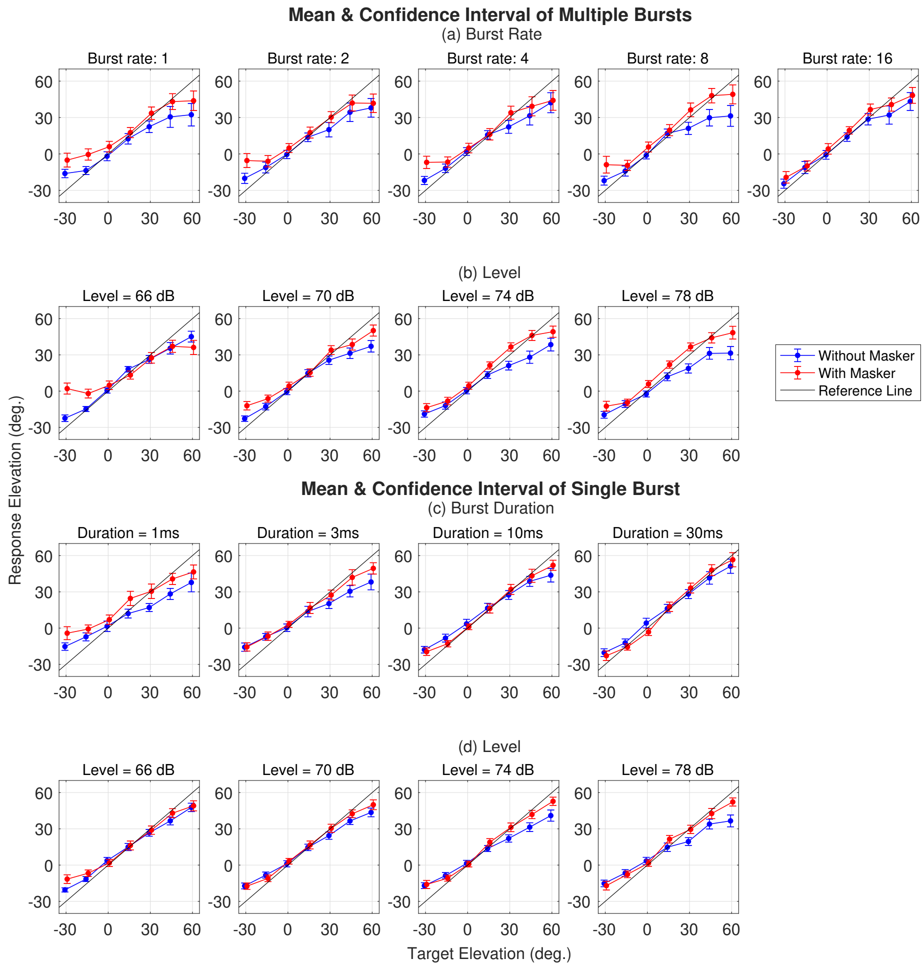

Figure 4: Mean values and 95% confidence intervals of the response data from both stimuli with and without the masker as a function of target source locations. The dots represent the mean values of each burst condition with and without the masker, respectively. The error bars are the confidence intervals of the means. The diagonal line represents the ideal fit. Rows a) and b) are the responses to each burst rate (level pooled) and level (burst rate pooled) for the multiple bursts, respectively. Rows c) and d) are the responses to each burst duration (level pooled) and level (duration pooled) for the single bursts.

was 60 dB SPL, the burst levels indicate the SNRs of 6, 10, 14, and 18 dB for the bursts with the masker. In most conditions, the 'With Masker' responses appear to be upward-biased towards a higher elevation than the 'Without Masker' responses. Moreover, the confidence intervals tend to be wider in the 'With Masker' results than those from the 'Without Masker' responses.

The mean responses of multiple-burst conditions show elevated localisation at around $-30°$ to $-15°$ and depressed localisation at around $30°$ to $45°$ of source elevation, which appear as s-shaped curves. This tendency can also be seen in single-burst conditions, but the curves appear more straight lines than s-shaped curves, for which the slope gains of the lines tend to be lower than the ideal fit (< 1). Since this tendency shows a narrower (compressed) degree range of the perceived elevation compared to the source positions' actual degree range, this localisation trend can be referred to as the *elevation compression effect.*

4.2 Elevation gains and biases

The elevation compression effect and the upward bias of the perceived elevation were quantified as a stimuli-response relation slope derived from a linear fitting. The present study designated the fitted line's slope and bias as the "elevation gain" and "elevation bias."

4.2.1 Elevation gains

Figures 5a, b, and c show the elevation gains from each stimulus condition, averaged over the elevation gains of each individual. The elevation gains from multiple-burst stimuli(Figure 5a) appear barely systematic, although there is a slight increase as the burst rate increases. The effect of the masker on elevation gain also appears to be insignificant. On the other hand, the results with the single-burst stimuli yield systematic effects both with and without the masker, which are clearly visible in Figure 5b. The elevation gain increases with the lengthening of the burst duration for both conditions with and without the masker. Notably, the elevation gains from stimuli with the masker appear to be generally higher than those from unmasked stimuli. The elevation gain of the stimuli with the masker increases above the unmasked bursts' elevation gain, approaching close to the ideal gain (unity) at 30 ms.

Figure 5c shows the elevation gains as a function of burst level. The masker noise appears to affect the mean elevation gain as the burst level increases. Whereas the mean elevation gain of stimuli without the masker decreases slightly as the level increases, the gain of stimuli with the masker rises above the elevation gain of unmasked bursts at 66–70 dB (6–10 dB SNR) and then levels off at 70–78 dB (10–18 dB SNR). In summary, the masker noise tended to improve the elevation gain as the duration of single-burst stimuli lengthened and the burst level increased.

4.2.2 Elevation biases

Figure 5d, e, and f show the elevation biases for each stimulus condition, averaged over the elevation biases of each individual. Generally, the stimuli with the masker show upward-biasing localisation compared to those without the masker, and elevation bias is barely affected by burst rate and level. However, it is noticeable that the

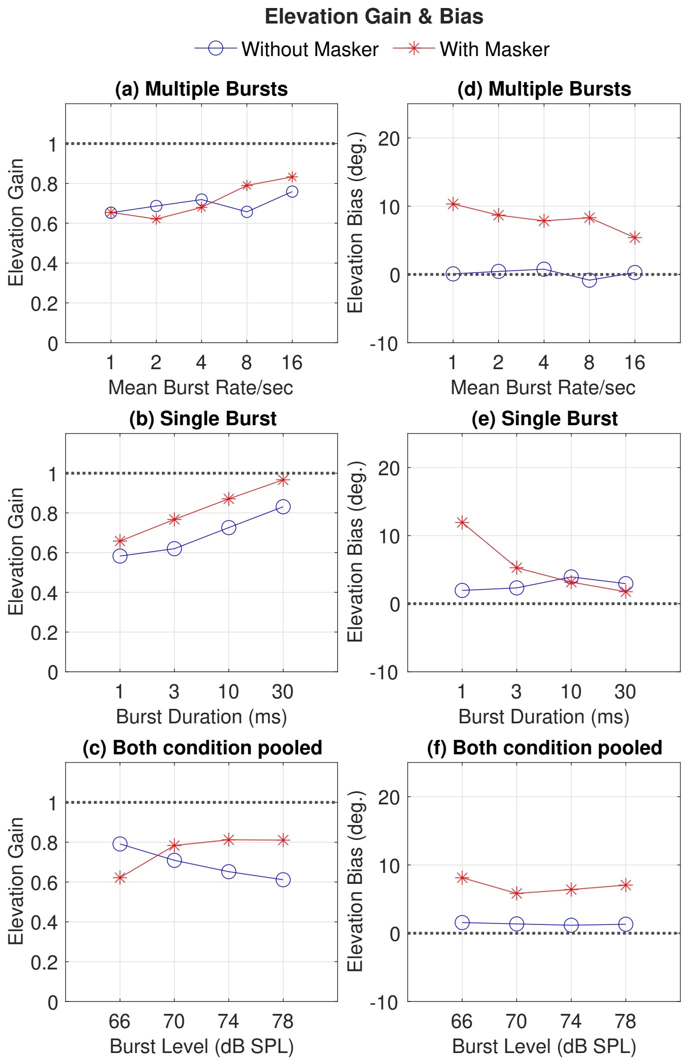

Figure 5: Elevation gain and bias from each stimulus condition. (a) The elevation gains of multiple bursts with and without masker as a function of mean burst rate between 1 and 16, where responses for each burst level were pooled. (b) The elevation gains of single bursts with and without masker as a function of burst duration between 1 ms and 30 ms, where responses for each burst level were pooled. (c) The elevation gains as a function of burst level with and without masker, where each condition for both multiple bursts and single bursts were pooled. (d) The elevation biases of multiple bursts with and without masker as a function of mean burst rate. (e) The elevation biases of single bursts with and without masker as a function of burst duration. (f) The elevation biases as a function of burst level with and without masker, where each condition for both multiple bursts and single bursts were pooled. The horizontal dotted lines in (a), (b), and (c) represent the ideal gain (unity) and the dotted lines in (d), (e), and (f) represent the unbiased condition.

masking noise significantly affects the single-burst stimuli, as shown in Figure 5e. While the single-burst stimuli without the masker scarcely yield a systematic effect on elevation bias by varying the burst duration, the elevation bias with the masker prominently decreases as the burst duration lengthens, approaching near 0° elevation at 30 ms.

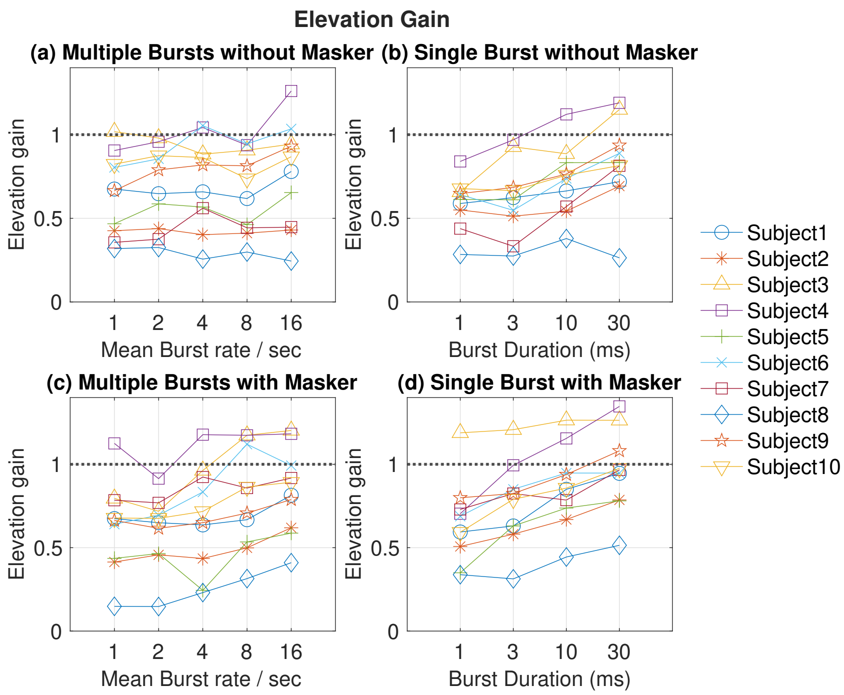

Figure 6: Elevation gain from every subject. (a) The elevation gains of "multiple bursts without the masker" conditions as a function of mean burst rate between 1 and 16. (b) The "single burst without the masker" conditions as a function of burst duration between 1 ms and 30 ms. (c) "Multiple bursts with the masker" conditions. (d) "Single burst with the masker" conditions. The horizontal dotted lines represent the ideal gain.

4.2.3 Individual variability

As Figure 5 only shows the mean values over the subjects' elevation gains and biases, it is necessary to examine the variability between each individual's results. Figures 6 and 7 show each subject's elevation gains and biases, highlighting significant diversity. Nevertheless, the overall trends depicted in Figure 5 generally correspond to the individual curves in specific conditions, such as elevation gains and biases from single burst stimuli. As the duration of the burst increases, most of each individual elevation gain increases for both single burst stimuli with and without the masker. Additionally, most of each individual elevation bias decreases significantly as the burst duration lengthens for single-burst stimuli with the masker.

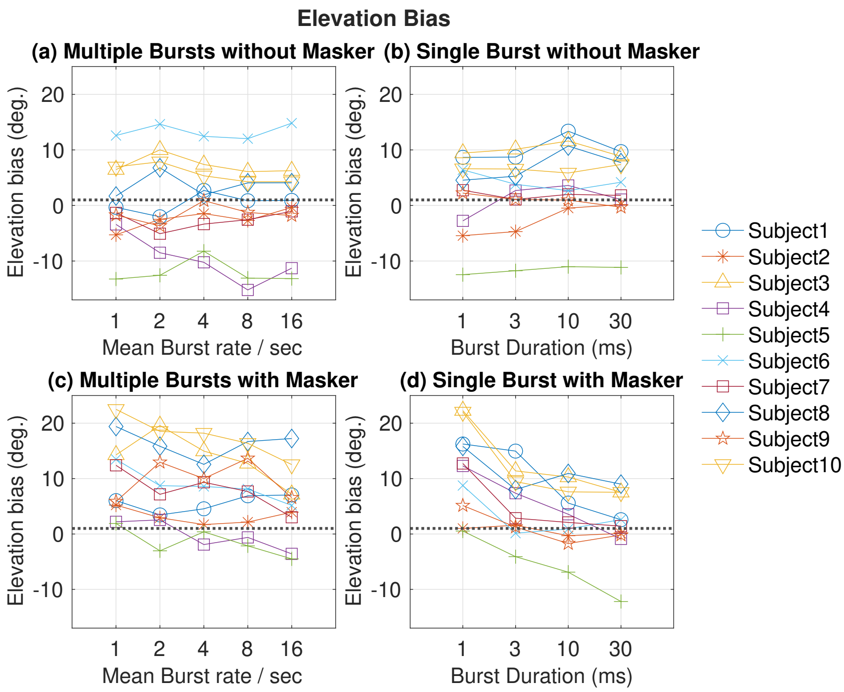

Figure 7: Elevation bias from every subject. (a) The elevation biases of "multiple bursts without the masker" conditions as a function of mean burst rate between 1 and 16. (b) The "single burst without the masker" conditions as a function of burst duration between 1 ms and 30 ms. (c) "Multiple bursts with the masker" conditions. (d) "Single burst with the masker" conditions. The horizontal dotted lines represent the unbiased condition.

However, without the masker, the burst duration has little effect on each individual elevation bias.

4.2.4 Statistical analyses

The statistical measures examining the trends displayed in Figure 5, 6, and 7 are presented in Table 1. Pearson's correlation coefficients measure the relationship between elevation gains or biases and the variations in each stimulus condition. The single-burst stimuli results (Burst Duration) show a moderate correlation between elevation gains and duration, with or without a masker. However, elevation gains of multiple-burst stimuli (Burst Rate) have weak correlations with burst rates. For the correlation with elevation biases, a high correlation between elevation biases and the duration of stimuli was observed with single-burst stimuli with the masker. However, all other conditions led to weak correlations with elevation biases.

The R^2 coefficient represents the goodness-of-fit measure for the regression models, which describes the strength of the relations between variations in elevation gains

Variable A	Variable B	Masking	Corr.	R^2	Adj.R^2	p value
Elevation Gain	Burst Rate	No	0.085	0.007	0.002	0.231
		Yes	0.221	0.049	0.044	**0.002**
	Burst Duration	**No**	0.366	0.134	0.129	**<.001**
		Yes	0.383	0.147	0.141	**<.001**
Elevation Bias	Burst Rate	No	-0.015	0	-0.005	0.832
		Yes	-0.175	0.031	0.026	**0.013**
	Burst Duration	No	0.074	0.005	-0.001	0.356
		Yes	-0.441	0.195	0.190	**<.001**

Table 1: Pearson's correlation coefficient, the coefficient of determination (R^2), and the adjusted R^2 values of the elevation gains and biases in each stimuli condition. Variables A and B show the dependent and independent variables (stimuli conditions), respectively. The cases with $p < 0.05$ are highlighted in bold characters in this table and for Tables 2–5

or elevation biases and each independent variable (burst rate and burst duration). Although R^2 coefficients generally show weakness in the measure due to individual variability, the single-burst stimuli exhibit higher coefficients with elevation gains and burst duration compared to others. In contrast, the elevation gains obtained from multiple-burst stimuli were far less consistent. In the elevation bias results, a relatively higher R^2 value was present only for the single-burst stimuli with masker.

4.3 Localisation errors, error and front-back confusion rates

4.3.1 *Mean signed differences of localisation errors*

Figure 8 shows the mean signed differences (MSD) and their 95% confidence intervals of the responses as a function of mean burst rates, burst duration, and levels of each multiple-burst and single-burst stimuli, both with and without the masker. These figures indicate the variability and dispersion in every test condition. The mean signed differences generally agree with elevation bias results (Figures 5d, e, and f) since the mean differences of the stimuli with the masker are more upward-biased than the unmasked stimuli in all conditions. Specifically, the upward bias appears more prominently with the multiple-burst stimuli. The signed differences for the multiple-burst and unmasked single-burst stimuli appeared barely affected by varying the mean burst rate or duration. However, the mean difference of the single-burst stimuli with the masker decreases as the burst duration lengthens (see Figure 8b).

4.3.2 *Error rates and front-back confusion rates*

The present study assessed the error rates by the responses outside the specific target ranges, which were ±15° and ±30° intervals referring from the target locations.

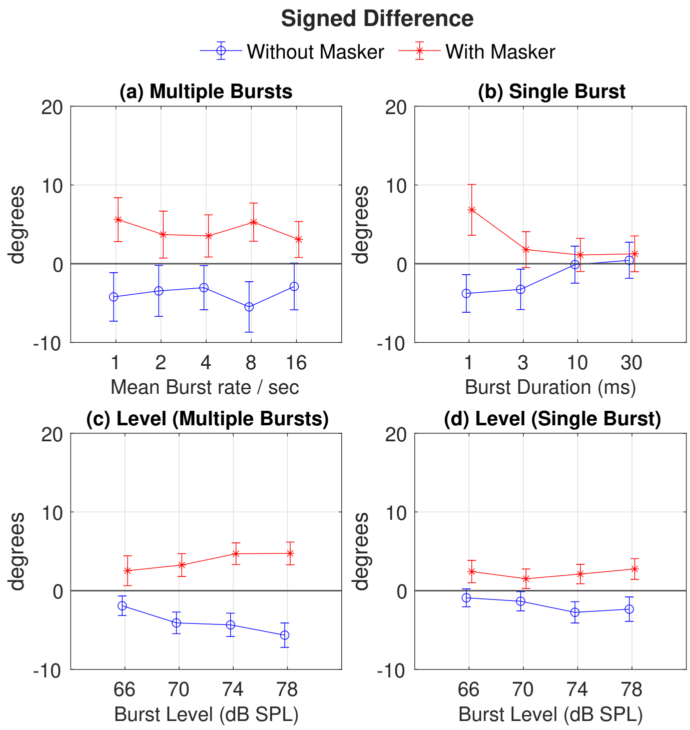

Figure 8: Mean signed differences and their 95% confidence intervals as a function of (a) burst rate (level pooled), (b) burst duration (level pooled), (c) level of multiple-burst stimuli (burst rate pooled), and (d) level of single-burst stimuli (burst duration pooled), for both conditions with and without the masker. Since the masker level was 60 dB SPL, the levels of the bursts with the masker indicate the SNRs of 6, 10, 14, and 18 dB.

Figure 9 shows the error rates derived from each target range and the front-back confusion rates as a function of the mean burst rate/sec, burst duration, burst levels, and target positions. The $\pm 15°$ ($30°$) intervals were determined considering the interquartile ranges of signed errors, which ranged between $20°$–$33°$, and the $\pm 30°$ intervals were added for a more comprehensive acceptance rate. Thus, error rates outside the $\pm 15°$ and $\pm 30°$ intervals represented the 'narrow-range' and 'wide-range' errors in the frontal hemifield, respectively. Front/back confusion rates were derived from the raw response data, which included all responses.

As seen in Figure 9, the error-rate plots show similar patterns as elevation-gain plots (Figure 5). Figure 9a shows that the error rates from multiple-burst stimuli appear barely systematic, although there is a slight decrease as the burst rate increases. The masker appears to increase the error rates of multiple-burst stimuli. However, the single-burst stimuli's error rates decline significantly for both target ranges with or without the masker as the duration increases (see Figure 9b), resembling the decreasing pattern shown in Figures 5e and 8b. Furthermore, the error rates are lower with masked stimuli than unmasked stimuli when duration$\geq$3 ms.

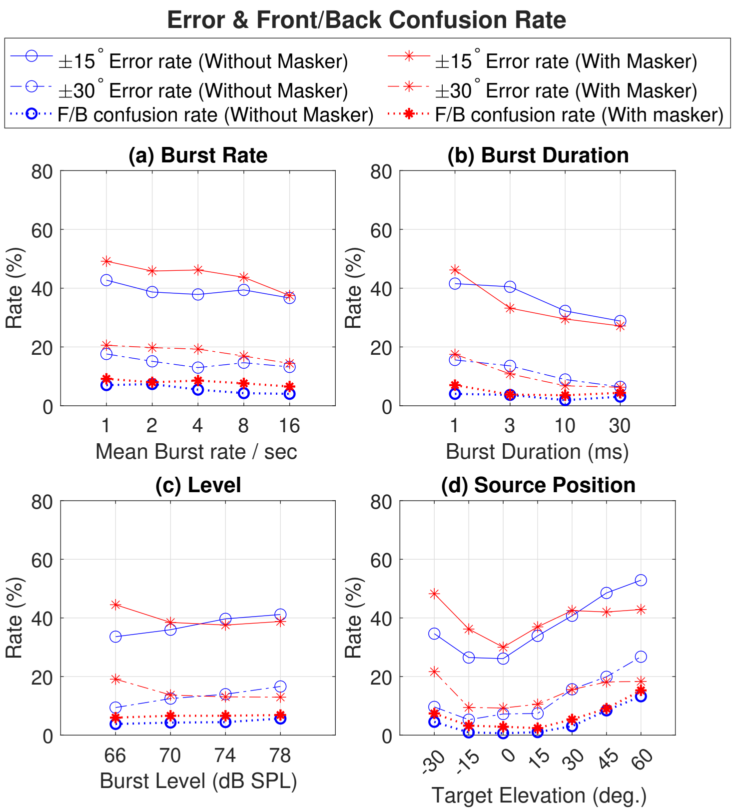

Figure 9: Error rates and front-back confusion rates as a function of mean burst rate, burst duration, burst levels, and target locations. The error rates were obtained from the response rates outside the ±15° and ±30° intervals centred from the target locations. The front/back confusion rates were obtained from raw data. (a) and (b): Error rates of multiple-burst and single-burst stimuli, respectively. (c) Error rates as a function of burst level from the combined data of multiple-burst and single-burst stimuli, where burst rates and duration are pooled. (d) Error rates as a function of the target location from the combined data, where burst rates, duration, and levels are pooled.

The improvement of localisation accuracy by the masker can also be observed in Figures 9c and 9d, showing that error rates are lower with masked stimuli than unmasked stimuli under certain conditions, such as the level of 74–78 dB (14–18 dB SNR), and 45°–60° elevation in the target location. Notably, as seen in Figure9c, the error rates of the masked stimuli decrease slightly, whereas those of unmasked stimuli increase as the stimuli level rises. These results imply that the masker tends to improve the localisation performance in these specific conditions. Besides, the front-back confusion rates barely show any significant correlation to the burst rates, duration, and levels despite the front-back confusion rates of masked stimuli being about 2–3% higher than the unmasked stimuli in all conditions.

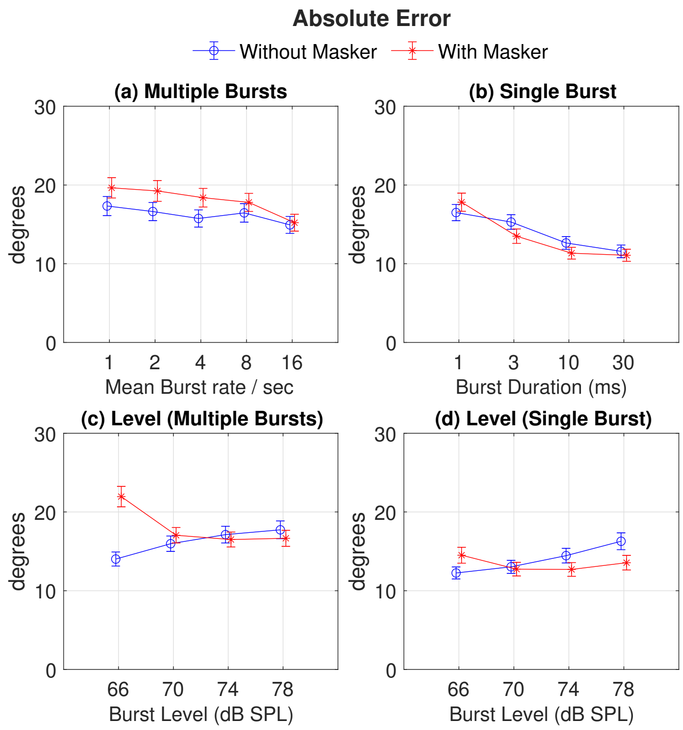

Figure 10: Mean absolute errors and their 95% confidence intervals as a function of mean burst rate, burst duration, and burst level for each multiple-burst and single-burst stimuli. (a) and (b): Absolute errors of multiple-burst and single-burst stimuli, respectively (stimuli levels are pooled). (c) Error rates as a function of burst level from multiple-burst stimuli (burst rates are pooled). (d) Error rates as a function of burst level from single-burst stimuli (burst duration is pooled).

4.3.3 Statistical analyses

To statistically examine the effect of the masker on localisation performance, the authors compared the responses' deviations from target locations (absolute errors) between unmasked stimuli and stimuli with the masker. Figures 10a and c show the mean absolute errors and their confidence intervals for multiple-burst stimuli as a function of burst rate and level, and Figures 10b and d show those of single-burst stimuli as a function of burst length and level. For the multiple-burst stimuli, the masker led to higher mean absolute errors at all burst rates (see Figure 10a) and burst levels up to 70 dB (see Figure 10c). However, the absolute errors of single-burst stimuli (Figure 10b and d) show that the masker caused slightly lower mean absolute errors for such conditions as 1) duration ≥ 3 ms and 2) level ≥ 70 dB (SNR ≥ 10 dB).

The statistical measures of the difference between the absolute errors were carried out using ANOVA and corresponding post-hoc tests. The results of absolute errors

Cases	Sum of Squares	df	Mean of Sqares	F	p
Rate	9831.417	4	2457.854	7.314	**< .001**
Level	2057.182	3	685.727	2.041	0.106
Rate x Level	9023.487	12	751.957	2.238	**0.008**
Masker	9120.679	1	9120.679	27.141	**< .001**
Rate x Masker	775.178	4	193.794	0.577	0.680
Level x Masker	26754.941	3	8918.314	26.539	**< .001**
Rate x Level x Masker	2281.623	12	190.135	0.566	0.871
Residuals	$2.586 \times 10^{+6}$	7694	336.046		

Table 2: ANOVA measures of the difference between absolute errors from multiple-burst stimuli responses in various conditions. Burst rate, level, and the presence of a masker are used as variables.

from multiple-burst and single-burst stimuli are shown in Table 2–3 and 4–5, respectively. The significance values (p) of the variables, including the term "Masker," show the statistical difference between unmasked stimuli and stimuli with the masker in various conditions. For the multiple-burst stimuli, no significant difference due to the masker has been measured in each burst rate (Rate * Masker), but a significant difference from the masker is found in the "Level * Masker" case. The corresponding post-hoc tests for the "Level * Masker" case (Table 3) found that, with the burst level of 66 dB, the masker significantly increased the absolute error but had no significant effect for the stimuli of higher levels, as apparent in Figures 10c.

Level	Mean Difference	SE	t	p
66 dB	−8.434	0.835	−10.106	**< .001**
70 dB	−1.455	0.835	−1.743	0.659
74 dB	0.175	0.831	0.211	1.000
78 dB	1.023	0.836	1.223	0.925

Table 3: Post-hoc test for the Level*Masker case of multiple-burst stimuli. Comparisons have been made between the absolute errors of multiple-burst stimulus pairs with and without a masker at each burst level.

Likewise, the results from the single-burst stimuli indicate no significant difference due to the masker in each burst duration, despite the apparent effect with duration $\geq$ 3 ms is shown in Figure 10b. The small sample size of the study (ten subjects) might result in wide variability in individual responses, potentially leading to limited statistical power in regression analyses and ANOVA results. However, a significant difference is found in the "Level * Masker" case for the single-burst stimuli. The corresponding post-hoc test found a noticeable decrease in absolute error at a burst level of 78 dB caused by background noise (see Figure 10d and Table 5). This result suggests that the enhancement in localisation performance due to the masker becomes evident at 78 dB or higher.

Cases	Sum of Squares	df	Mean of Squares	F	p
Duration	39990.571	3	13330.190	51.686	**< .001**
Level	2242.789	3	747.596	2.899	**0.034**
Duration x Level	1508.106	9	167.567	0.650	0.755
Masker	1883.581	1	1883.581	7.303	**0.007**
Duration x Masker	1494.971	3	498.324	1.932	0.122
Level x Masker	5437.478	3	1812.493	7.028	**< .001**
Duration x Level x Masker	4816.654	9	535.184	2.075	**0.028**
Residuals	$1.620 \times 10^{+6}$	6282	257.908		

Table 4: ANOVA measures of the difference between absolute errors from single-burst stimuli responses in various conditions. Burst duration, level, and the presence of a masker are used as variables.

Level	Mean Difference	SE	t	p
66 dB	−1.823	0.801	−2.277	0.307
70 dB	1.163	0.806	1.442	0.837
74 dB	1.761	0.813	2.165	0.373
78 dB	3.270	0.815	4.014	**0.002**

Table 5: Post-hoc test for the Level*Masker case of single-burst stimuli. Comparisons have been made between the absolute errors of single-burst stimulus pairs with and without a masker at each burst level.

5 DISCUSSION

The present experiment investigated the effect of added broadband noise on the median plane localisation of short noise bursts. The results for single noise burst mainly agree with earlier studies on the effects of duration and level of short sounds on median plane localisation [18, 27, 38]. However, since the influence of added noise is extensively investigated in this study, the results show some noticeable findings related to the presence of a broadband masker. The localisation of aperiodic multiple noise bursts was also examined, but the effect of the average burst rate was barely systematic. This section discusses the multiple-noise-burst localisation first, and the findings from the single-noise bursts are discussed in the following parts.

5.1 Localisation of aperiodic repetition of short noise bursts

Multiple occurrences of 1 ms long noise bursts were presented with random temporal distribution to examine the localisation performance of short sounds in the median plane. As a result, the average burst rate of aperiodic noise bursts provided little systematic influence on the elevation gain and bias, as seen in Figure 5 and Table 1. This result is in conflict with the responses from Hofman and Van Opstal's experiment [18] conducted with periodic burst-train stimuli, where the elevation gain increased with the repetition rate (i.e., narrowing the silence gap between bursts). The stable elevation perception achieved with "periodic" repetition of short bursts is likely due to the auditory system accumulating the directional information periodically.

However, the present experiment only resulted in minor decreases in the error rates by varying the burst rate, as is evident in Figure 9. This difference might be because 1) the burst duration was 1 ms whereas Hofman and Van Opstal (1998) [18] employed 3-ms-long bursts, 2) the noise bursts were aperiodically presented, and 3) the burst rate was adjusted for the listeners to detect every single burst. Thus, the limited effect of multiple noise bursts can be initially interpreted as the spectral information being limited due to the insufficient burst duration of each single burst. Nonetheless, accumulating the spectrum of short bursts was also deficient for the auditory system to compensate for the insufficient directional information because of low burst rates and aperiodic repetition.

In Hofman and Van Opstal's study, the elevation gain declines significantly to lower values (0.6–0.7) when the stimuli duration decreases to 40–80 ms. Since the silence gap at the highest burst rate of 16 (i.e., the minimum average silence gap) in the present experiment varies from a minimum of 10 ms to a maximum of 115 ms, there might be a chance that the auditory memory of recent directional information fades away during the aperiodic burst presentation. Consequently, directly comparing the present result with Hofman and Van Opstal's study might be inappropriate, and treating multiple noise bursts as a random presentation of a single 1 ms burst would be feasible.

Despite the minor influence of burst rate on the elevation gain and error rate, mean elevation biases of unmasked multiple-burst stimuli were about 2°–4° lower than for the unmasked single-burst stimuli (see Figure 5d and e). This result might be a minor effect observed from the aperiodic "repetition", where every single burst of the stimuli was detectable.

5.2 Elevation compression effect

As mentioned earlier, the elevation compression effect refers to the compression of the perceived elevation range compared to the sources' actual angle range, and it manifests itself as elevation gains less than unity. This effect is observed in most test conditions, as seen in Figures 4 and 5. However, Figure 5b shows rises in elevation gain with increasing stimulus duration, rising to the mean elevation gain of around 0.7 already with 10-ms stimuli for both masked and unmasked bursts. This finding supports the neural integration theory suggested by previous studies such as [18] and [38]. In contrast to our study, Hofman and Van Opstal's outcome [18] for the stimulus duration securing a stable elevation perception was 80 ms, where the elevation gain was between 0.6 and 1, whereas our experiment, despite individual variability, resulted in a much shorter stimulus duration (10 ms) for a stable localisation.

One hypothesis for the elevation compression effect is that the precision in localising a sound source may be weakened when the sound source is located away from the centre. Previous studies have reported that localisation errors and variability increase as the source location rises in the median plane [8, 28], whose results mostly agree with the present study's results.

Besides, in previously mentioned studies, the elevation compression effect was observed with sound stimuli of relatively long durations, considering this phenomenon as a consistent effect for the median plane localisation. Nevertheless, a few studies [e.g. 18, 38] show that the elevation gain varies with different durations of sound stimuli. For example, the elevation decreases by shortening the duration, which might be associated with the lack of perceived spectral detail due to the short duration of the sound stimulus. This lack of spectral detail could consequently result in precision weakness. This hypothesis is supported by the auditory analysis of the macroscopic spectral cues studied by Kim et al. (2022) [21], who report that the spectral variability of the head-related transfer function (HRTF) 'envelope' is reduced as the elevation rises, which could lead to a reduction in localisation precision.

Figure 11 shows the difference between HRTF envelopes of adjacent target locations, where the envelopes are derived by averaging the individual HRTFs in an HRTF library [5]. These HRTF difference envelopes, namely *the difference spectra* [21], indicate the macroscopic changes, which lack spectral details, between generic HRTFs of the median plane locations from 0° to 80°. Thus, spectral variation in these changes can imply localisation sensitivity between locations. While the upper plots (0°–10° and 10°–20°) show many variations across a broad spectral range, the curve gradually flattens, excluding a narrow spectral region at 8–10 kHz, as the HRTF position rises. The plots through 40°–50° to 70°–80° show only slight variability, indicating that almost negligible macroscopic spectral changes emerge between 40° and 80° elevation. The variability is even smaller for the locations above 80° according to Kim et al. (2022) [21]. These difference curves suggest that the localisation precision weakens as the direction of the target sound source rises from 0° to 80° elevation.

Considering that these difference spectra are smoothed and thus lack spectral details such as elevation-dependent spectral notches, short sound bursts perceived with insufficient spectral detail would also be increasingly mislocalised in the median plane since the sound source is located away from 0° elevation. Sufficient spectral detail is achieved when the sound stimulus is sustained long enough, thus providing the listeners with higher localisation accuracy and elevation gain.

5.3 Effect of background noise on elevation bias

The responses of multiple-burst stimuli with the masker appear to be generally upward biased than unmasked stimuli, as shown in Figures 4 and 5d. One can hypothesise that upward-biasing localisation can be a discrimination process in the auditory system due to auditory masking, namely the pushing effect. A few previous

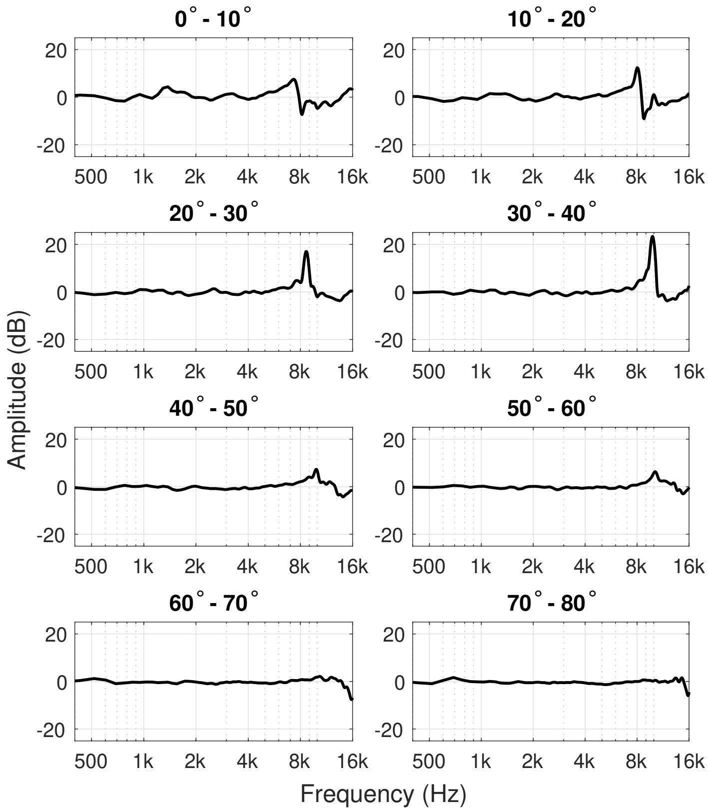

Figure 11: Difference between macroscopic HRTF envelopes of adjacent target locations. The adjacent target location angles are marked in the title of each graph. For example, the difference spectrum titled 0-10° is the difference between the HRTF envelope measured at 10° elevation and that measured at 0°. The HRTF envelope curves are derived by averaging HRTFs from the *HUTUBS* library [5].

studies [7, 20] examined the pushing effect on auditory localisation when noise was presented before the target sound. They suggested that the auditory system first adapts to the location of a prior noise. A target sound, when presented after a prior sound, is perceived as being shifted further away from the prior sound's location to help improve detecting the latter sound, especially when spectral information of both sounds coincides. This hypothesis could also be applied to explain the localisation of the present experiments since background noise lasting at least 300 ms was present before the target stimulus.

However, the pushing effect cannot explain the upward bias in the target location below the masker located at $\phi = 0°$ because stimuli at those locations were *pulled* towards the masker. The reason for the upward bias effect due to the masking is unknown. Still, assuming that the discrimination process in the auditory system involves more than just shifting the perceived location away from the masker location is possible. Instead, the upward bias may involve biasing the inner localisation coordinate system itself.

The elevation bias due to the single-burst stimuli is noticeable. The biases of the masked stimuli decrease significantly as the stimulus duration is lengthened (see Figures 5e and 7d), whereas stimuli without the masker do not correlate much with the stimulus duration. This decreasing tendency shows a relatively higher correlation and determination level than other conditions, as seen in Table 1. This phenomenon implies that elevation bias is effectively reduced by a noise masker while the duration of the stimulus is lengthened, especially when a spectrally similar masker is presented before the target stimuli.

5.4 Effects of background noise on localisation performance with different sound levels

Figure 5c shows mean elevation gain as a function of sound pressure level. The elevation gain from unmasked bursts declines consistently from 66 dB to 78 dB SPL with both multiple-burst and single-burst stimuli. The decreasing trend of elevation gain agrees with the results obtained by Vilegan and Van Opstal (2004) [38], where the elevation gain gradually declines above 55–65 dB SPL. Likewise, the error rates for the unmasked bursts also show an increasing trend with increasing levels (see Figure 8). However, these results only partly agree with the *negative level effect* found in [17], where the error rate abruptly rises from 80 dB SPL and above. A negative level effect seems to occur even at lower sound levels in the present study, since the moderate increase in error rate is seen in unmasked bursts ranging from 66 dB to 78 dB SPL. Still, this moderate increase below 80 dB is also apparent in Hartmann and Rakerd's study [17].

On the contrary, the mean elevation gain curve for masked bursts shows an increasing trend. The gain rises from 66 dB to 70 dB (6 dB to 10 dB SNR) and levels off from 70 dB to 78 dB (10 dB to 18 dB SNR). This pattern is similar to the result in Macpherson and Middlebrooks (2000) [27] presented with background noise. Moreover, the mean elevation gains of masked bursts rise even higher than those of unmasked bursts from around 70 dB (10 dB SNR), as evident in Figure 5c. Masking appears to lessen the negative level effect and further enhance localisation performance. The effects of the masker noise observed in the present study agree with the peripheral adaptation theory presented by Macpherson and Middlebrooks (2000) [27], suggesting that a masker helps the auditory system to pre-adapt to a high-level stimulus within 10–20 dB SNR, thus improving the localisation performance.

The effect of masking on the elevation gain and error rate of single-burst stimuli is noteworthy. The specific conditions where the improvement of elevation gain is statistically significant are found to be SNR $\geq$ 18 dB ($\geq$ 78 dB), as seen in Figure 10 and Table 5. The improved localisation performance in this condition is also revealed in the error rates. The masking noise reduced error rates by 3.7% in the 'wide range' and 2.4% in the 'narrow range.' errors.

5.5 Influence of implicit factors and prior belief

Besides the hypotheses discussed in this study, it is worth noting that a group of studies have supported the elevation compression effect through the use of Bayesian inference, which incorporates directional features and prior beliefs to estimate auditory localisation [e.g., 2, 12, 19]. Listeners' prior beliefs indicate their accumulated experience and knowledge of the environment, such as their long-term exposure to directional information in everyday life. Since the directions of sound sources are perceived inequitably, incorporating prior beliefs into the distribution of perceived directional information would reveal biases in auditory localisation.

For example, Hofman and Van Opstal (2002) [19] reported "compression" in the response elevation range between $-30°$ and $30°$, Barumerli et al. (2023) [2] found elevation responses biased toward the horizontal plane, and Ege et al. (2018) [12] reported increasing precision (lower response variability) in the localisation of the frontal area at the cost of a decrease in accuracy (elevation gain). Therefore, it is essential to take into account that the study's finding of a bias towards the frontal direction in localisation could be affected by various factors such as the influence of the oculomotor range (the range of human eyesight), the spatial distribution statistics of sound sources in the environment [31], or the listeners' previous knowledge and post-experience about the test conditions, such as perceived target range [11].

6 CONCLUSION

Auditory localisation of short broadband noise burst(s) was examined in the median plane with and without spectrally similar masking noise. The noise burst comprised multiple bursts lasting 1 ms, aperiodically repeated with an average burst rate between 1 and 16 times per second, and single-burst stimuli of duration between 1 ms and 30 ms. They were presented with (and without) incoherent spectrally similar masker noises generated at $\pm30°$ azimuth in the horizontal plane. Localisation performance was investigated for different average burst rates, burst duration, sound levels (SNRs between 6 dB and 18 dB), the presence of the masker on the elevation gain, elevation bias, signed and absolute errors, and error rate. The results are summarised below.

1. The localisation of aperiodically repeated multiple-burst stimuli was found to have a weak systematic effect on the elevation gain and bias. Likewise, an increase in burst rate led to a minor decrease in the localisation error rate. When background noise was added, the responses appeared more upward biased than unmasked stimuli and showed higher absolute errors and error rates than unmasked stimuli for all burst rates.

2. The elevation compression effect is observed in most test conditions for single-burst stimuli, represented by lower elevation gain than unity. The mean elevation gain increased as the stimulus duration increased, which reached around 0.7 already with a 10-ms-long stimulus with or without background noise.

3. The mean elevation bias of single-burst stimuli with background noise was higher than unmasked stimuli at burst durations of 1–3 ms. However, the elevation bias decreased significantly as the stimulus duration was lengthened. Despite individual variability, the elevation bias for single-burst stimuli with the masker correlated moderately with the burst duration.

4. The gain and bias curves exhibited substantial individual variability. Nonetheless, the individual gain and bias curves for single-burst stimuli generally corresponded to the mean curves.

5. The error rate of single-burst stimuli declined significantly as the stimulus duration increased. In certain conditions where the duration was at least 3 ms, error rates were lower for stimuli with background noise than unmasked stimuli. This result suggests that added noise can improve localisation performance in this specific condition.

6. For unmasked stimuli, the mean elevation gain decreased gradually, and the error rates increased with increasing the sound level. However, with background noise, the elevation gain curves showed an increasing trend, and the error rates appeared to decrease as the sound level increased. The mean elevation gains for stimuli with noise appeared even higher than those for unmasked stimuli from around 70 dB (10 dB SNR). Accordingly, the background noise tended to restrain the deterioration of localisation at high sound levels and further improve localisation performance.

7. The absolute errors between stimuli with and without a masker showed no statistically significant difference from multiple-burst stimuli at each burst rate and from single-burst stimuli at each burst duration. However, when the signal-to-noise ratio in single-burst stimuli was at least 18 dB, the background noise caused a statistically significant reduction in the mean absolute error.

References

[1] Asano, F., Suzuki, Y., and Sone, T. Role of spectral cues in median plane localization. *Journal of the Acoustical Society of America 88* (1990), 159–168.

[2] Barumerli, R., Majdak, P., Geronazzo, M., Meijer, D., Avanzini, F., and Baumgartner, R. A bayesian model for human directional localization of broadband static sound sources. *Acta Acustica 7* (5 2023), 12.

[3] Best, V., Schaik, A. V., Jin, C., and Carlile, S. Auditory spatial perception with sources overlapping in frequency and time. *Acta Acustica united with Acustica 91* (2 2005), 421–428.

[4] Blauert, J. *Spatial Hearing: The Psychophysics of Human Sound Localization.* MIT Press, 1997.

[5] Brinkmann, F., Dinakaran, M., Pelzer, R., Grosche, P., Voss, D., and Weinzierl, S. A cross-evaluated database of measured and simulated hrtfs including 3d head meshes, anthropometric features, and headphone impulse responses. *J. Audio Eng. Soc 67*, 9 (2019), 705–718.

[6] Brungart, D. S., and Simpson, B. D. Effects of bandwidth on auditory localization with a noise masker. *Citation: The Journal of the Acoustical Society of America 126* (2009), 3199.

[7] Canévet, G., and Meunier, S. Effect of adaptation on auditory localization and lateralization. *Acta Acustica united with Acustica 82*, 1 (1996), 149–157.

[8] Carlile, S., Leong, P., and Hyams, S. The nature and distribution of errors in sound localization by human listeners. *Hearing Research 114* (12 1997), 179–196.

[9] Cornsweet, T. N. The staircase-method in psychophysics. *The American Journal of Psychology 75*, 3 (1962), 485–491.

[10] Dizon, R. M., and Litovsky, R. Y. Localization dominance in the median-sagittal plane: Effect of stimulus duration. *The Journal of the Acoustical Society of America 115* (6 2004), 3142–3155.

[11] Ege, R., Opstal, A. J. V., and Wanrooij, M. M. V. Perceived target range shapes human sound-localization behavior. *eNeuro 6* (3 2019).

[12] Ege, R., Van Opstal, A. J., and Van Wanrooij, M. M. Accuracy-precision trade-off in human sound localisation. *Scientific Reports 8* (11 2018), 16399.

[13] Fastl, H., and Zwicker, E. *Psychoacoustics.* Springer Berlin Heidelberg, 2007.

[14] Good, M. D., and Gilkey, R. H. Auditory localization in noise. i. the effects of signal[U+2010]to[U+2010]noise ratio. *The Journal of the Acoustical Society of America 95* (5 1994), 2896–2896.

[15] Good, M. D., and Gilkey, R. H. Auditory localization in noise. ii. the effects of masker location. *The Journal of the Acoustical Society of America 95* (5 1994), 2896–2896.

[16] Goupell, M. J., Majdak, P., and Laback, B. Median-plane sound localization as a function of the number of spectral channels using a channel vocoder. *The Journal of the Acoustical Society of America 127* (2 2010), 990–1001.

[17] Hartmann, W. M., and Rakerd, B. Auditory spectral discrimination and the localization of clicks in the sagittal plane. *The Journal of the Acoustical Society of America 94* (10 1993), 2083–2092.

[18] Hofman, P. M., and Van Opstal, A. J. Spectro-temporal factors in two-dimensional human sound localization. *The Journal of the Acoustical Society of America 103* (5 1998), 2634–2648.

[19] Hofman, P. M., and Van Opstal, A. J. Bayesian reconstruction of sound localization cues from responses to random spectra. *Biological Cybernetics 86* (4 2002), 305–316.

[20] Kashino, M., and Nishida, S. Adaptation in the processing of interaural time differences revealed by the auditory localization aftereffect. *The Journal of the Acoustical Society of America 103* (6 1998), 3597–3604.

[21] Kim, T., Pöntynen, H., and Pulkki, V. Vertical direction control using difference-spectrum filters in stereophonic loudspeaker reproduction. *Journal of the Audio Engineering Society 70* (3 2022), 128–139.

[22] Kulkarni, A., and Colburn, H. S. Role of spectral detail in sound-source localization. *Nature 396* (12 1998), 747–749.

[23] Langendijk, E. H., and Bronkhorst, A. W. The contribution of spectral cues to human sound localization. *The Journal of the Acoustical Society of America 105* (1999), 1036–1036.

[24] Langendijk, E. H. A., Kistler, D. J., and Wightman, F. L. Sound localization in the presence of one or two distracters. *The Journal of the Acoustical Society of America 109* (5 2001), 2123–2134.

[25] Levitt, H. Transformed up[U+2010]down methods in psychoacoustics. *The Journal of the Acoustical Society of America 49* (1971), 467–477.

[26] Lorenzi, C., Gatehouse, S., and Lever, C. Sound localization in noise in normal-hearing listeners. *The Journal of the Acoustical Society of America 105* (3 1999), 1810–1820.

[27] Macpherson, E. A., and Middlebrooks, J. C. Localization of brief sounds: Effects of level and background noise. *The Journal of the Acoustical Society of America 108* (10 2000), 1834–1849.

[28] Makous, J. C., and Middlebrooks, J. C. Two-dimensional sound localization by human listeners. *Journal of the Acoustical Society of America 87* (1990), 2188–2200.

[29] Moore, B. C. J. *An introduction to the psychology of hearing*, 6 ed. Emerald Group Publishing Limited, 2012.

[30] Morimoto, M., and Ando, Y. On the simulation of sound localization. *Journal of the Acoustical Society of Japan (E) 1*, 3 (1980), 167–174.

[31] Parise, C. V., Knorre, K., and Ernst, M. O. Natural auditory scene statistics shapes human spatial hearing. *Proceedings of the National Academy of Sciences of the United States of America 111* (4 2014), 6104–6108.

[32] Perrett, S., and Noble, W. The effect of head rotations on vertical plane sound localization. *The Journal of the Acoustical Society of America 102* (1997), 2325–2332.

[33] Pulkki, V., Delikaris-Manias, S., and Politis, A. *Parametric time-frequency domain spatial audio.* Wiley Online Library, 2018.

[34] Pulkki, V., and Karjalainen, M. *Communication Acoustics: An introduction to speech, audio and psychoacustics.* Wiley, 2015.

[35] Pulkki, V., Pöntynen, H., and Santala, O. Spatial perception of sound source distribution in the median plane. *Journal of the Audio Engineering Society 67* (11 2019), 855–870.

[36] Shelton, B. R., and Scarrow, I. Two-alternative versus three-alternative procedures for threshold estimation. *Perception & Psychophysics 35* (1984), 385–392.

[37] Thurlow, W. R., and Runge, P. S. Effect of induced head movements on localization of direction of sounds. *Citation: The Journal of the Acoustical Society of America 42* (1967), 480.

[38] Vliegen, J., and Van Opstal, A. J. The influence of duration and level on human sound localization. *The Journal of the Acoustical Society of America 115* (4 2004), 1705–1713.

[39] Wightman, F. L., and Kistler, D. J. Headphone simulation of free-field listening. ii: Psychophysical validation. *Journal of the Acoustical Society of America 85* (1989), 868–878.

[40] Wightman, F. L., and Kistler, D. J. Resolution of front–back ambiguity in spatial hearing by listener and source movement. *http://dx.doi.org/10.1121/1.426899* (4 1999).

[41] Zwiers, M. P., Van Opstal, A. J., and Cruysberg, J. R. M. A spatial hearing deficit in early-blind humans. *The Journal of Neuroscience 21* (5 2001), RC142–RC142.

A Responses for every condition

Figures 12 and 13 show violin plots as a function of source direction, displaying response data for every test condition, including average burst rate (Figure 12), burst duration (Figure 13), sound pressure level, and the masker noise's presence or absence.

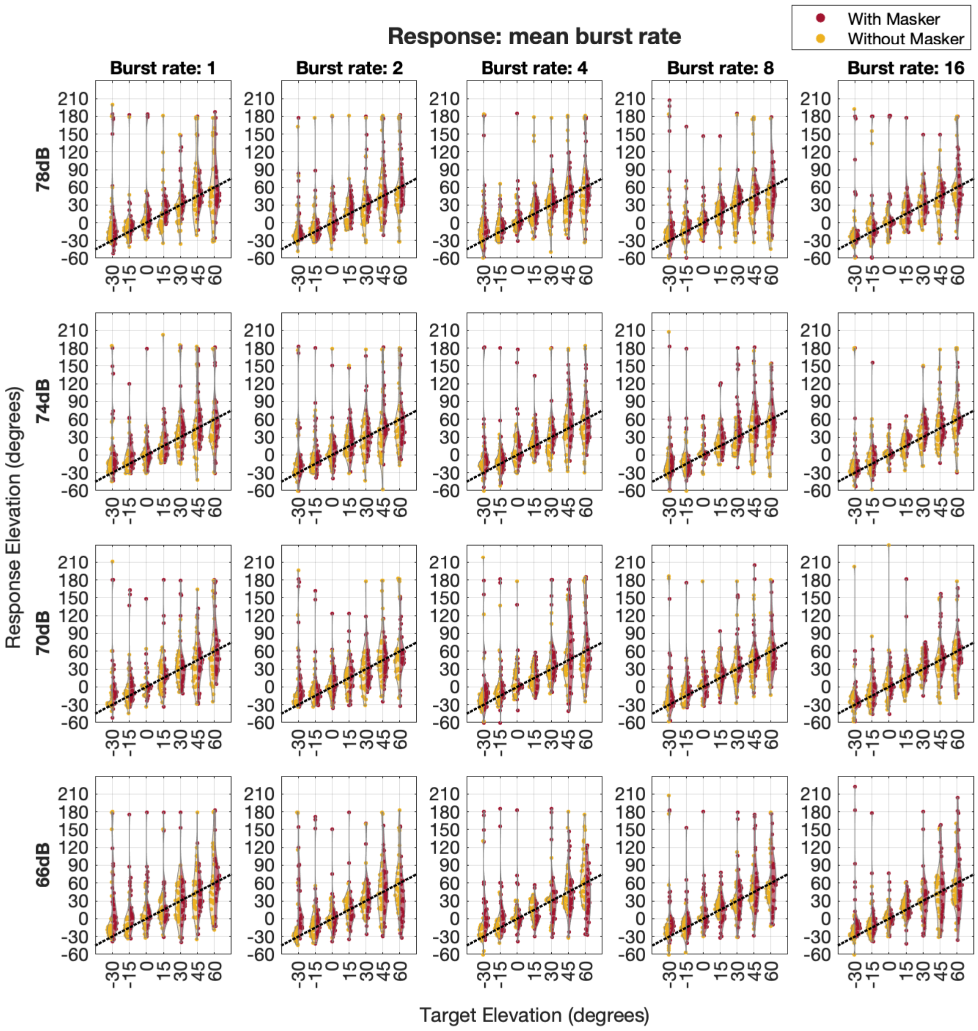

Figure 12: The violin plot as a function of source location for multiple-burst stimuli. The left-side plots indicate the responses to unmasked stimuli, and the right-side plots indicate the responses to masked stimuli. The straight line presents the ideal stimulus-response relation.

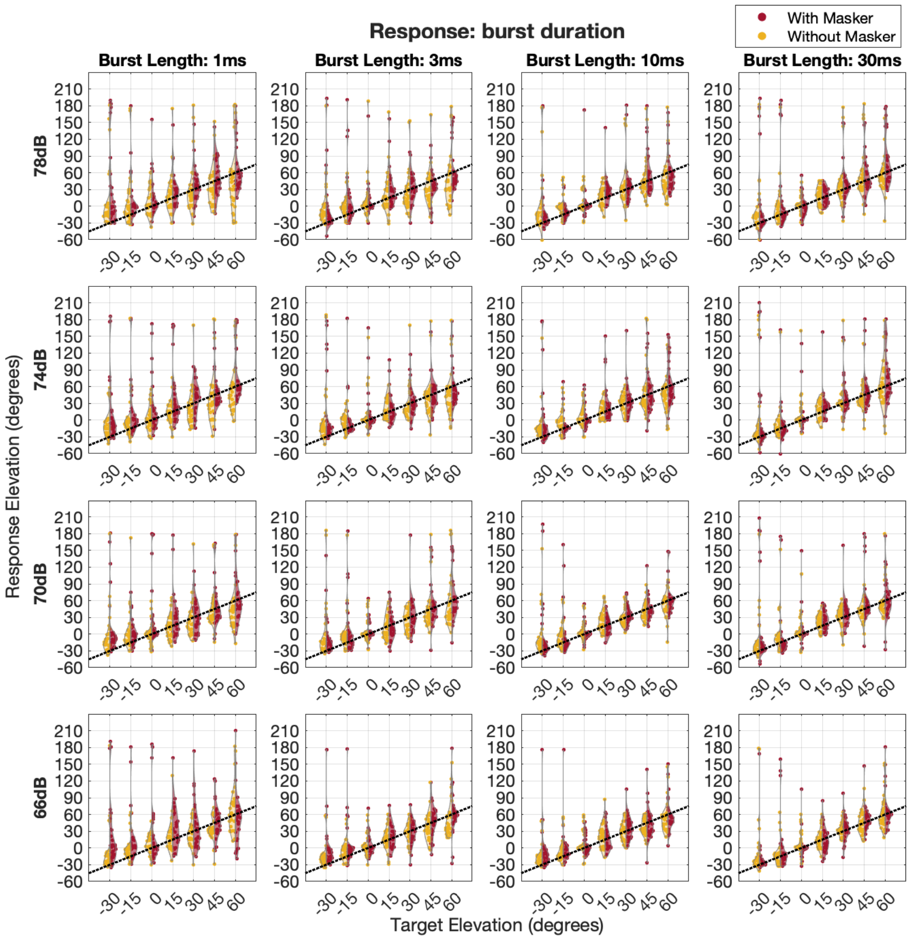

Figure 13: The violin plot as a function of source location for single-burst stimuli. The left-side plots indicate the responses to unmasked stimuli, and the right-side plots indicate the responses to masked stimuli. The straight line presents the ideal stimulus-response relation.

B Elevation gains and biases from each stimulus condition.

Figures 14 and 15 show elevation gains for each burst level as a function of burst rate and elevation biases for each burst level as a function of burst duration, with and without the masker.

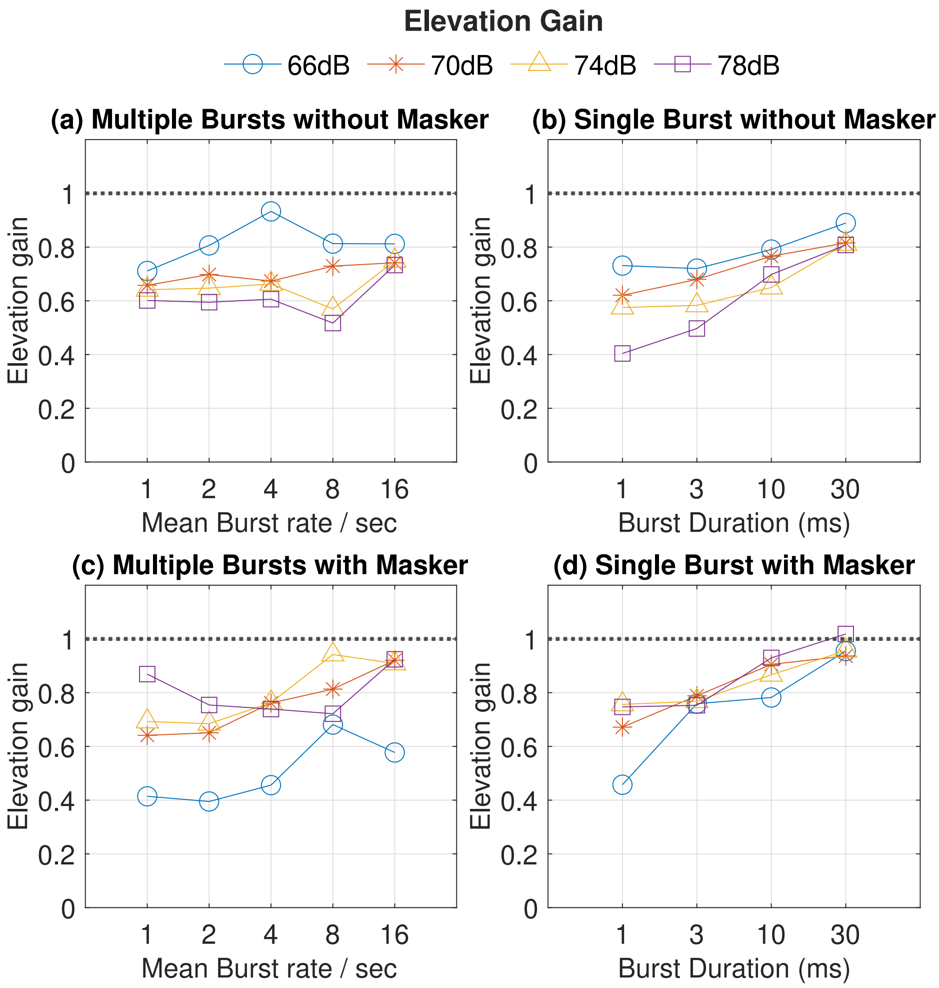

Figure 14: Elevation gain from each stimulus condition. (a) The elevation gains of “multiple bursts without the masker” conditions as a function of mean burst rate between 1 and 16. (b) The “single burst without the masker” conditions as a function of burst duration between 1 ms and 30 ms. (c) “Multiple bursts with the masker” conditions. (d) “Single burst with the masker” conditions. The horizontal dotted lines represent the ideal gain (unity).

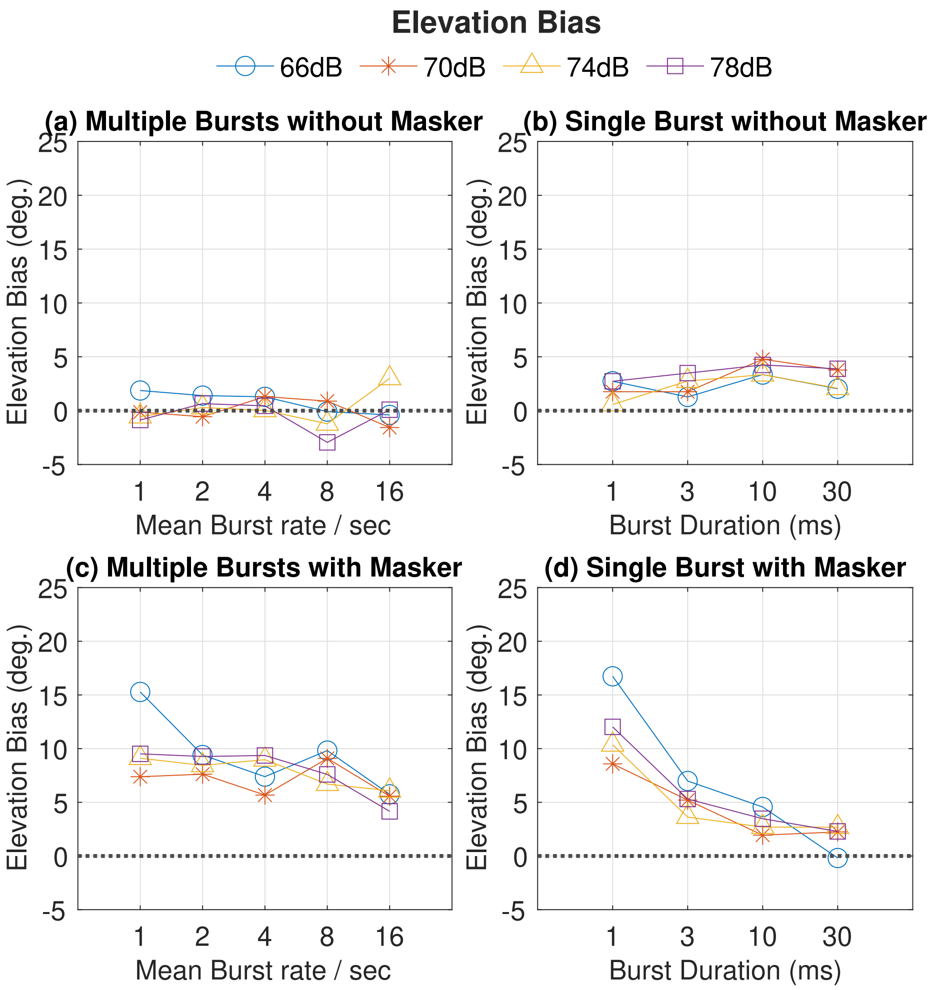

Figure 15: Elevation bias from each stimulus condition. (a) The elevation biases of "multiple bursts without the masker" conditions as a function of mean burst rate between 1 and 16. (b) The "single burst without the masker" conditions as a function of burst duration between 1 ms and 30 ms. (c) "Multiple bursts with the masker" conditions. (d) "Single burst with the masker" conditions. The horizontal dotted lines represent the unbiased condition.