Design and Evaluation of Personal Audio Systems based on Speech Privacy Constraints

Daniel Wallace and Jordan Cheer
Institute of Sound and Vibration Research, University of Southampton, UK [a)]

Personal Audio refers to the generation of spatially distinct sound zones that allow individuals within a shared space to listen to their own audio material without affecting, or being affected, by others. Recent interest in such systems has focussed on their performance in public spaces where speech privacy is desirable. To achieve this goal, speech is focussed towards the target listener and a masking signal is focussed into the area where the target speech signal could otherwise be overheard. An effective masking signal must substantially reduce the intelligibility in this region without becoming an annoyance to those nearby. To assess these perceptual requirements, listening tests were carried out using two examples of loudspeaker arrays with different spatial aliasing characteristics, to determine the impacts of different masking signal spectra on speech intelligibility and subjective preference. The results of these tests were used, alongside objective and subjective metrics, to form a design specification for private personal audio systems.

 [http://dx.doi.org(DOI number)]

[XYZ] Pages: 1–13

I. INTRODUCTION

In spaces shared by multiple people, sound produced for the attention of one listener may become annoying or distracting to others nearby. This situation worsens as the number of competing sources and listeners increases, motivating the provision of *personal sound zones*[1] using loudspeaker arrays. The technical requirements of these arrays have been discussed in various contexts, such as open plan offices and museum exhibits[1], television and radio systems[2–4], entertainment devices for use in vehicles and aircraft[5,6] and mobile devices[7]. In each application, systems are configured to focus an input signal into a *bright zone*, where the target listener is located, whilst minimising the leakage of the signal into the *dark zone*. System performance has conventionally been measured in terms of the *acoustic contrast*[8], i.e. the sound level difference between zones, but trade-offs between this indicator and subjective measures of performance have been identified[9,10]. Consequently, recent research has begun to incorporate elements of psychoacoustics into the design of loudspeaker arrays[2,11] and sound zoning algorithms[12–14] to improve the subjective performance of personal audio systems, making the technology more appropriate to the solution of real-world problems. Following this approach, the present paper integrates the results from listening tests and perceptual metrics into the personal audio system design process.

One application where perceptual requirements influence the design process is when a personal audio system must provide *speech privacy control*. The ability of a personal audio system to control the level of sound in different spatial regions can be leveraged to privately transmit a spoken message to an individual in a public space, as first investigated by Donley et. al.[15]. This objective is achieved when the message is intelligible within the bright zone of the system and is rendered unintelligible in the dark zone, where it could otherwise be overheard. In many envisaged applications of this control scheme, there exists a defined region where intelligibility reduction must be prioritised. In an in-vehicle telecommunication system[16], for example, the locations of other listeners in the vehicle are known. Alternatively, such a system may be used to transfer speech through a glass security partition at a bank or post office counter; here, the greatest potential for eavesdropping occurs at the front of the queue, or at adjacent cashier locations. In either case, leakage of spoken information from the bright zone into the dark zone could represent a loss of privacy for the target listener. Limitations in the acoustic contrast achievable by conventional systems means that speech leaked into the dark zone could remain intelligible; to mitigate this, a pair of sound field control processes may be employed[15]. The first process focuses a speech signal towards the target listener in the bright zone, and the second uses the same array to focus a secondary masking signal into the dark zone, with the purpose of impairing the intelligibility of any leaked speech. The performance of such a system can therefore be characterised by the speech intelligibility contrast (SIC)[15] i.e. the difference between the intelligibility of the spoken message in the bright and dark zones, evaluated using an objective speech intelligibility metric.

As noted by Donley et. al[17], one effect that can lead to a degradation in the SIC is spatial aliasing in the array. Depending on the source and zone geometries, a dramatic reduction in the privacy performance can occur as

[a)] D.Wallace@soton.ac.uk, J.Cheer@soton.ac.uk;

side lobes in the array directivity at high-frequencies may lead to signal leakage between the zones. The standard geometrical argument states that at frequencies greater than $0.5c/\delta$, where δ is the inter-element spacing and c is the speed of sound, spatially aliased side-lobes will be present in the array directivity. The work of Donley et. al.[17] provides an analytic derivation for the frequency at which spatially aliased side-lobes from one beam impinge on the opposite zone, with the recommendation that this frequency be used to band-limit the masking signal. These conclusions were based on free field simulations, with validation through anechoic chamber measurements, and close matching was found between the two[17]. However, no jury testing was carried out to assess the objective or subjective system performance, rather, instrumental metrics were used for this purpose. Furthermore, the anechoic testing only offers limited insight into the system performance in a realistic room, even with a modest reverberation time. In such spaces, energisation of the reverberant field by spatially aliased side-lobes will further decrease acoustic contrast, even if they do not directly intercept the opposite zone. Earlier work by Donley et. al.[15] showed results from three simulated reverberant rooms, with reverberation times of less than 0.5 seconds, using a 3 metre diameter, 295-channel circular loudspeaker array. While adequate SIC performance was demonstrated, such a system is clearly difficult to realise.

Previous work by the present authors has discussed a number of practical considerations required for the implementation of personal audio systems for speech privacy control. An assessment of the experience of all listeners in a space, rather than just the target listener, was first discussed in terms of masking signal design[18]. Analysis based on the Psychoacoustic Annoyance metric[19] predicted that although speech-shaped maskers are most effective at providing SIC, filtering out high frequencies could provide acceptable speech privacy whilst also reducing the potential for the masker to be annoying or distracting to nearby listeners, by reducing its *sharpness*. Accordingly, in the present work the masking performance and overall subjective preference of various low-pass filtered speech-shaped maskers are assessed by listeners.

Also considered in prior work was the case where part of the masking effect was provided by ambient background noise in the space[20], but this masking contribution is difficult to predict in practice due to fluctuations in the background noise level. As such, the systems presented in this work are assumed to be operating in quiet, or at least be the dominant acoustic sources in a space, such that background noise can be neglected in calculations of the signal-to-noise ratio (SNR).

The purpose of the present paper is to demonstrate a method for designing a speech privacy control system based on a combination of subjective metrics and physical array limitations. A system in this regard comprises the physical layout of the loudspeakers and the masking signal emitted by them. The results of objective and subjective listening tests are used to provide greater insight into the design of personal audio systems for speech privacy enhancement over the previous research in this area, which have relied directly on instrumental measures of speech intelligibility and quality.

Section II describes the geometry of the two arrays, whose sizes have been chosen to enable an investigation into the effect of spatial aliasing on masking signal design. In Section III, the technical capability of these two array configurations is discussed in terms of acoustic contrast, followed by a general discussion of how this relates to the control of speech privacy in Section IV. The design of objective and subjective listening tests to evaluate the performance of these systems is presented in Section V, and the results from these tests are compared against objective and subjective metrics in Sections VI and VII. A summary of the proposed design method is presented in Section VIII and conclusions follow in Section IX.

Throughout this paper, subscripts $\{\}_b$ and $\{\}_d$ refer to quantities related to the bright and dark zones respectively. The two loudspeaker arrays used for the purposes of investigating the dependence on spatial aliasing are referred to as *narrow* and *wide* arrays with corresponding subscripts $\{\}_n$ and $\{\}_w$.

II. LOUDSPEAKER ARRAY GEOMETRY

In order to investigate the design of a personal audio system for speech privacy, including the effects of spatial aliasing, a physical loudspeaker array is used as a testbed. This 27-channel array was originally designed for in-car 3D audio reproduction to two listeners simultaneously[21]. The effects of loudspeaker spacing are investigated by selecting two sub-arrays, each with $L = 9$ elements, from the full array. Figure 1 shows a front-view of the geometry of the full array with the two sub-arrays highlighted. The alternating vertical offset of odd and even drivers allows the horizontal driver spacing δ to be minimised. The *narrow* array, indicated by solid outlines in Figure 1, has loudspeakers spaced at $\delta_n = 0.035$ m intervals and the *wide* array (dotted outlines) has $\delta_w = 3\delta_n = 0.105$ m, yielding array widths of $D_n = 0.28$ m and $D_w = 0.84$ m. The array is positioned at a height of 1.2 metres, at the centre of a listening room, which has a mid-frequency reverberation time of 0.11 s, and dimensions 4.4 m × 3.7 m × 2.2 m.

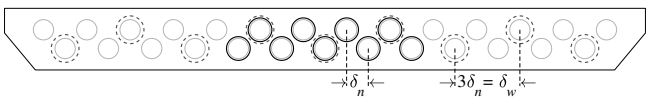

FIG. 1. Front view of loudspeaker array. Groups of 9 elements were selected from a 27-channel array to form two arrays with different horizontal element spacing. Narrow and wide sub-arrays are indicated with solid and dotted lines respectively.

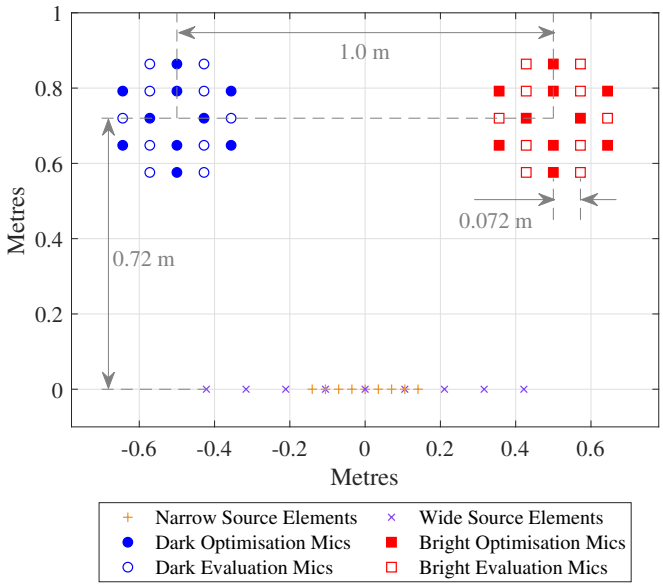

FIG. 2. [Colour Online] Plan view of the personal audio system geometry, showing source and microphone locations.

Figure 2 shows a plan view of the personal audio systems in the listening room. The bright and dark zones are specified by the positions of two microphone arrays. In both cases, the centres of the bright and dark zones are situated 0.72 m in front of the loudspeaker array in order for all microphones to be within the critical distance[22] of the loudspeaker array in the room ($\approx$ 1 m). Outside of the critical distance, the sound field is dominated by diffuse, reflected sound and this would significantly impede sound field control. The centres of the two zones are spaced 1.0 m apart, which is comparable to the physical width of the wide array, and is thus consistent with its intended performance limitations. In the interest of generality, this symmetrical geometry is chosen to match the de facto standard for evaluating personal audio systems that has emerged from the literature, e.g.[10,12,14,17,23]. Nevertheless, of the applications mentioned above, the zonal layout is a close match to the in-car reproduction scenario, when considering providing separation between the left and right sides of the car cabin. Each zone has a radius of 0.15 m, covering enough space for a human head, with 20 microphones distributed on a grid within each zone. Half of the microphones are used to optimise the zoning filters, while the other half are used to evaluate the reproduced sound field. Disjoint sets of microphones are used to reduce bias in the acoustic contrast estimates that are presented in the following section[23,24].

III. OBJECTIVE PERFORMANCE EVALUATION

Before considering the subjective performance of the narrow and wide arrays, it is important to understand their objective performance in terms of acoustic contrast, as the physical limitations of the two configurations directly affect how the signals reproduced by them are perceived. For each array configuration, two parallel acoustic contrast control (ACC) processes are used to focus the speech and masker into their respective zones. ACC is chosen as, compared to other sound zoning methods, it will by definition of the optimisation procedure offer the greatest level of contrast between zones when used in a reflective environment[23]. This is important, since the speech-to-masker ratio is highly correlated with intelligibility, and this is largely governed by the acoustic contrast between the zones. According to the ACC method, the loudspeaker weights at each frequency, $\mathbf{q}_b$, that focus the speech programme into the bright zone are found by determining the eigenvector corresponding to the largest eigenvalue of

$$[\mathbf{G}_d^H\mathbf{G}_d + \beta\mathbf{I}]^{-1}[\mathbf{G}_b^H\mathbf{G}_b] \quad (1)$$

where $\mathbf{I}$ is an $L \times L$ identity matrix, $\mathbf{G}_b$ and $\mathbf{G}_d$ are measured electroacoustical transfer responses from the array sources to the bright and dark zone optimisation microphones respectively, and β is a regularisation parameter. At each frequency, β is set to be proportional to the condition number of $\mathbf{G}_d^H\mathbf{G}_d$, with proportionality constant $\beta_0 = 10^{-13}$. This value was selected to provide a trade-off between robustness to changes in the environment, acceptable acoustic contrast across the speech frequency range, and flatness of the frequency response. A 1/3 octave band equaliser is applied to the input signals to maintain the spectrum of the speech signal in the bright zone and the masker in the dark zone, compensating for any residual colouration to the frequency response caused by the ACC filters. The corresponding process to find $\mathbf{q}_d$ is identical, with $\mathbf{G}_b$ and $\mathbf{G}_d$ interchanged. The symmetrical zonal geometry yields the same acoustic contrast results for both ACC processes, but this is not the case in general; the leakage of the programme into the dark zone and of the masker into the bright zone must be considered separately, based on the contrast provided by each

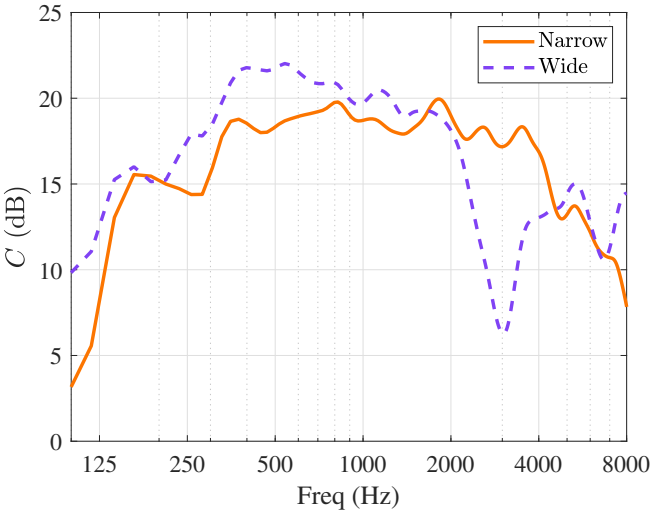

FIG. 3. [Colour Online] Acoustic contrast measurements for the narrow and wide loudspeaker array configurations.

process. Figure 3 shows the acoustic contrast, C, for the narrow and wide arrays. From these results it can be seen that the wide array has a slightly higher contrast at low-mid frequencies due to its wider aperture, but spatial aliasing due to the inter-element spacing causes a substantial reduction in contrast between 2 and 4 kHz.

To provide further insight into the acoustic contrast results presented in Figure 3, room impulse responses were captured using the microphone array grid depicted in Figure 2, positioned at multiple locations within the room. Output from the array was simulated using the weights $\mathbf{q}_b$ calculated above to produce maps of the radiated tonal sound fields; these results are presented in Figure 4 at 1.2, 2.4 and 4.8 kHz for each array configuration.

From these results it can be seen that at 1.2 kHz, the aperture of the wide array provides a more focussed beam pattern than the narrow array, while at higher frequencies, the aliasing limit of the wide array begins to become evident. With the presented zonal geometry, at 2.4 kHz a secondary lobe in the directivity begins to impinge on the dark zone; this explains the pronounced decrease in acoustic contrast around this frequency, as shown in Figure 3. At 4.8 kHz, the narrow array creates a tightly focussed beam in the direction of the bright zone, whereas the sound field in the room with the wide array is essentially diffuse, with multiple side lobes being radiated in different directions. In rooms with longer reverberation times, the ratio of direct to diffuse sound around the loudspeaker array will decrease, hindering sound field control at a distance from the array. Furthermore, individual reflections may impinge on the dark zone in the same way as is demonstrated with aliased side-lobes in Figure 4[25]. Undertaking a sound field mapping exercise such as the one provided here, or a ray-tracing simulation with important reflections included, gives significantly more information to system designers than predictions of inter-zone contrast alone, at the cost of increased computational or measurement complexity. Figures 3 and 4 illustrate the physical limitations of the two array configurations. However, neither acoustic contrast measurements nor predictions of the spatial distribution of sound pressure provide a definitive indication of whether each system fulfils its intended purpose of providing speech privacy.

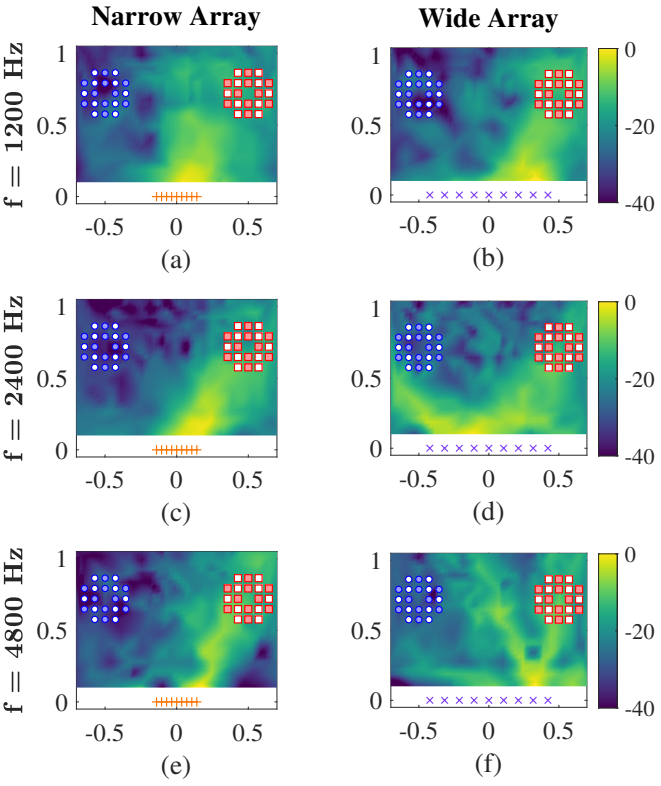

FIG. 4. [Colour Online] Relative Sound Pressure Level with tonal signals focussed to the bright zone (square markers) at 1.2, 2.4 and 4.8 kHz for each source array configuration. Dimensions are in metres.

IV. SUBJECTIVE ASPECTS OF SPEECH PRIVACY CONTROL

The privacy of a particular listener is violated when others can understand content that is intended to remain confidential. As discussed in the introduction, this is achieved in speech privacy control systems by both focussing the speech towards the target listener and radiating additional masking noise into the dark zone, where listeners who would otherwise be capable of eavesdropping are situated. In addition to pursuing privacy, successful systems must also remain considerate towards good-intentioned listeners who happen to be occupying the dark zone. These goals, being judged subjectively by humans rather than measured, significantly complicate the process of designing such a system. Therefore, the first step in making the design process tractable is to identify proxies for privacy and the subjective preference of listeners.

A. Speech Privacy

Privacy is intimately linked with speech intelligibility, a quantity which itself can be measured in many different ways. Intelligibility is often defined as the percentage of sounds, words, keywords or sentences that are correctly identified during speech-in-noise listening tests[26–28]. The knowledge gained from such listening tests and studies has led to the development of a range of objective intelligibility prediction methods for different applications. One such measure, which has been linked to the assessment of speech privacy, is the Speech Intelligibility Index (SII) and its predecessor, the Articulation Index (AI)[26], whose values can be interchanged for AI<0.5 using an empirically derived relationship[29].

A conceivable design framework for a speech privacy control system, therefore, is to predict the SII in each zone from a standard speech spectrum and the known frequency response of the array. This can then be used

to adjust the masker to satisfy a minimum SII in the bright zone, and a maximum SII in the dark zone. This improves upon the objective of maximising SIC, used by Donley et. al[17], as it explicitly defines the intelligibility requirements in each zone. The remaining question is then how to set these requirements.

Subjective experiments on office privacy by Cavanaugh et. al.[30] reported that for speech transmission through office walls, the break point between adequate and inadequate "confidential privacy" occurs at an equivalent SII of 0.10, and at SII=0.17, everyday privacy requirements are met. These claims have been validated in more recent studies of office privacy[29] and are referenced (in terms of the AI) in the current ASTM standard for objectively measuring speech privacy in open plan spaces[31]. Other research has related SII values between 0.22 and 0.33 to the Speech Reception Threshold (SRT) of 50% words correct in listening tests with short, meaningful sentences[32]. This range of SII values suggests that for low values of the SII, small changes in the SII result in large differences in the level of intelligibility[33]. This feature of the SII led Gover and Bradley[34] to reject the use of the metric as a measure of speech *security*, which is a more stringent condition encompassing the audibility of speech sounds and vocal cadence, as opposed to the intelligibility of individual words. In the same work, 50% intelligibility of words within sentences was reported at SII = 0.11, and the threshold for understanding at least one word corresponded to an SII of 0.05[34]. Leakage of the masker into the bright zone can impede intelligibility for the target listener, so complementary *minimum* SII limits must be set in this region. The standard that defines the SII states that good communication systems have an SII in excess of 0.75 and poor communication systems have an SII below 0.45[35].

The wide range of SII values described above can be attributed to the broad scope of the referenced experiments and nuances in the various definitions of privacy used in each instance. Since none of the aforementioned experimental contexts exactly align with the specific task of creating private sound zones, independent listening tests will be carried out in order to define the requirements of the zonal privacy system.

B. Masker Preference

In addition to providing objective information about the intelligibility of speech radiated from the array, listeners can also provide important subjective information regarding their preference for different masking signals. These preference results can then be correlated with the properties of the original stimuli to determine which metrics most accurately predict listener preference in the considered context. One potential metric, Psychoacoustic Annoyance[19,36], has previously been used to predict listeners' negative reaction to noise in a variety of contexts, for example in university facilities[37], to rate household fans[38] and traffic noise[39]. *Psychoacoustic* Annoyance refers to a component of the overall sensation of annoyance that can be directly attributed to the acoustical structure of a noise. It does not take into account other contributions to the actual annoyance that cannot be obtained through signal analysis alone, such as the perceived source of the sound, environmental context, or the time of day. The Psychoacoustic Annoyance metric, $A(x)$, is defined for a single input signal x, and is formulated as

$$A(x) = N_5(x)\left(1 + \sqrt{w_S^2(x) + w_{FR}^2(x)}\right), \qquad (2)$$

where

$$w_S(x) = (S(x) - 1.75) \times 0.25\log(N_5(x) + 10) \qquad (3)$$

for $S > 1.75$ acum, and

$$w_{FR}(x) = 2.18/N_5^{0.4}(x)(0.4F(x) + 0.6R(x)) \qquad (4)$$

where N_5 is the instantaneous loudness in *sones* exceeded for 5 percent of the input signal duration. Sharpness S is measured in units of *acum*, and is correlated with the level of high frequency content in the signal. Fluctuation Strength F and Roughness R are measured in units of *vacil* and *asper* respectively, and each report the level of low frequency ($\lesssim$ 20 Hz) and high frequency ($\gtrsim$ 20 Hz) modulations respectively.

The fluctuation of multi-talker babble usually results in more effective masking than speech-shaped noise reproduced at the same level[40], but is also reported to negatively affect the Psychoacoustic Annoyance rating. The question of whether babble is less annoying in practice is a topic for further investigation, but is likely to depend on the context, and the level and composition of ambient noise in the playback environment[20]. As a compromise, therefore, to avoid tying results to any particular context, stationary random noise maskers are considered in this work. This is also consistent with studies that reference the SII in the context of privacy control[29,32], and early studies into validating the Psychoacoustic Annoyance metric[36]. The perception of roughness in stationary band-limited random signals is caused by random amplitude fluctuations, but the sensation is greatly reduced for wideband signals compared to narrowband noise. Accordingly, from Equation 2 it can thus be hypothesised that reducing the sharpness and loudness of the masker are the predominant means by which a random noise masking signal can be reduced in psychoacoustic annoyance. The objective of reducing sharpness provides an alternative criterion for selecting the cut-off frequency for a low-pass filter applied to the masker, and can be compared objectively and subjectively against low-pass filtering the masker to prevent spatially aliased side-lobes impinging on the bright zone.

V. LISTENING TEST DESIGN

The success of a personal audio system for speech privacy control relies on its ability to control the level of in-

telligibility in each listening zone, without producing excess noise pollution. While instrumental metrics for these factors have been suggested, the values of these metrics at which adequate privacy performance is achieved are unclear from the literature. Accordingly, sentence intelligibility and paired preference tests were conducted, with listeners situated in the dark zone of the zonal audio system. Six array-masker configurations were tested in total, corresponding to two array widths with three cut-off frequencies for the low-pass filter applied to the masker. These conditions will be referred to using the code $[W|N]_f$, for the wide and narrow arrays respectively. The first pair of conditions, W_A and N_A, refers to the case where the cut-off frequency is set to the point where a spatially aliased side-lobe begins to impinge on the opposite sound zone, as illustrated for the wide array in Figure 4d. This cut-off frequency is 2.4 kHz for W_A and 8 kHz for N_A. For the second pair of conditions, W_S and N_S, the cut-off frequency is set to 4 kHz for both arrays, in order to filter out frequencies that contribute strongly to the sensation of sharpness[19]. The final pair of conditions, W_∞ and N_∞, refer to the case where the low-pass filter is bypassed.

A. Speech Intelligibility Test

The chosen speech intelligibility test is a modified version of the English Matrix Test[41,42], a descendent of the Swedish[43] and Oldenburg[44] matrix tests, so called as sentence material is built up from a "matrix" of individual words, forming unpredictable, but grammatically correct five-word sentences. The left panel of Figure 5 shows a screen-capture from the custom designed MATLAB interface, showing the ten options for each of the five words in each sentence. A closed-set presentation format, where the list of candidate words are visible to participants, was chosen to reduce training effects[42]. Additionally, the test induces less listener fatigue because the closed-set format slightly eases word understanding; SRTs are approximately 1 dB higher in open-set presentations of the same test[45]. Pilot testing confirmed that at the SRT of the closed-set test, participants reported that speech could be considered private, and that without access to the word list, understanding would be significantly impeded. This is consistent with the published slope of the reference psychometric function of 13%/dB SNR at the SRT[45], i.e. a 1 dB change in SNR results in a change of 13% in the intelligibility score.

The stimuli were presented to the test subjects using an auralisation process. Sentences in noise at the desired SNR were convolved with the acoustic contrast control filters and binaural room impulse responses measured from the loudspeaker array to the ears of a KEMAR mannequin situated in the dark zone, facing the centre of the array, and were presented to listeners over open-backed headphones in a soundproofed room. 20 sentences were presented for each condition, and the SNR was adjusted adaptively[46] based on the percentage of correct words at each presentation, relative to the target score

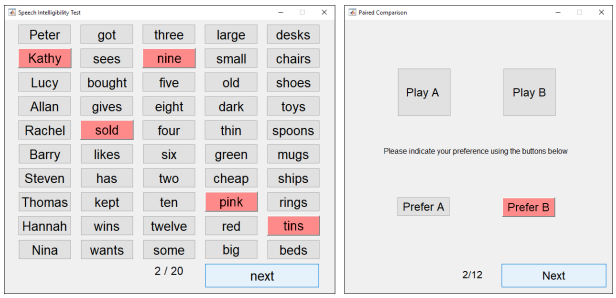

FIG. 5. [Colour Online] Screen captures from the matrix test interface after a sentence has been presented (left) and the preference test interface (right).

of 50%. One training list of 20 sentences was presented before the first sentence test, with a higher target intelligibility score of 70%, so that participants could become familiar with the speaker's voice and the response format.

B. Preference Test

To assess which of the three low-pass filter configurations was preferred for each array width, stimuli were sampled from the intelligibility test at the SNR that corresponded with the SRT, and were aggregated into a paired comparison test. For each participant, the overall level of the three samples in each test were adjusted based on pilot studies to ensure that all participants were comparing similar noise levels, without affecting the SNR. The right panel of Figure 5 shows the interface for the preference test. The instruction to participants was to listen to both unlabelled stimuli, then, focussing only on the noise, rather than any speech that may be perceivable in the background, select which of the two noises they prefer. Participants were asked to make a simple preference judgement, rather than rate or rank the sounds in terms of their annoyance or any other attribute, due to the potential for participants having different internal definitions of these quantities. Twelve comparisons were made for each array width, which constituted 4 repeats of 3 pairs of cut-off frequencies.

$N = 21$ participants from across the University of Southampton were invited to take part in the study. All were between 18 and 40 years of age ($\mu = 26.1, \sigma = 3.7$ years), had self-reported normal hearing and were fluent in the English language. In total, the test took around 45 minutes for each participant to complete. Participants volunteered their time for the study. The following two sections discuss the results of the sentence and preference tests in turn, comparing each set of results with objective and subjective metrics.

VI. SENTENCE TEST RESULTS AND DISCUSSION

The results of the sentence test are SRTs in dB, and are presented in Figure 6 as a series of box plots for each array and masker configuration. The SRTs for the wide

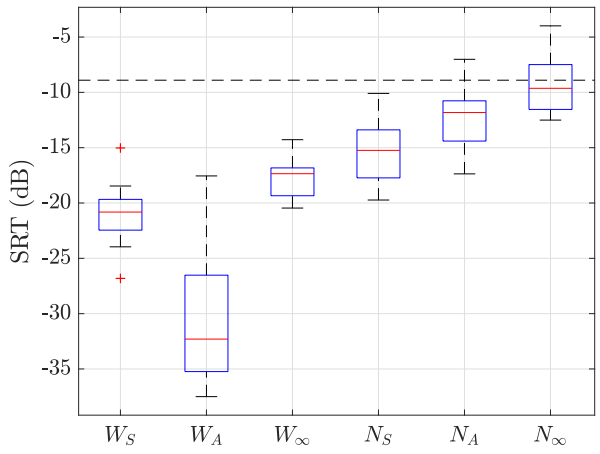

FIG. 6. [Colour Online] Distribution of Speech Reception Thresholds (SRTs) achieved for each array and masker configuration. N=21 participants. Red plusses indicate outliers, identified as the results which lie greater than 1.5 times the box length from the edges of the box (approximately $\pm 2.7\sigma$). The reference SRT for the closed-set English matrix test[42] is represented with a dashed horizontal line.

array are significantly lower than for the narrow array, indicating that the level of the masking signal must be increased to achieve the same level of (un)intelligibility. The wide array has poor high-frequency control due to spatial aliasing, so a significant amount of high frequency speech information is leaked into the dark zone. This necessitates an increase in masking signal level, compared to the narrow array, which has more consistent levels of acoustic contrast control in the speech frequency range, as shown in Figure 3. The variability of the acoustic contrast level across frequency is also responsible for the large range of SRTs recorded for condition W_A. The low-pass filtered masker in this condition cannot adequately mask consonant sounds, giving listeners increased opportunity to correctly guess words from the provided matrix of sentences. This enlarges the inter-subject variability beyond that usually expected from the normal-hearing population.

A. Comparison with SII

System designers require a target signal to noise ratio that corresponds to conditions of sufficient unintelligibility in the dark zone, SNR_d. The results from the listening test above demonstrate that SNR_d varies with different arrays and masking signals. The SII metric can be used to provide a single number value that predicts SNR_d, for a given array and masker configuration. To obtain this value, each stimulus that was presented to participants during the matrix test was also passed through the SII algorithm, providing an objective rating of intelligibility that could be compared with the listening test score for that sentence. Each participant listened to a set of 20 sentences in each test condition. For each of these sets, the percentage of correct words identified in each stimulus was fitted to the corresponding SII using a sigmoid function. Each curve of SII values against intelligibility was then interpolated at the 50% words correct level, and when this intermediate value is averaged across all participants and conditions, a score of 50% words correct in the matrix test corresponds with an SII value of 0.05. This value is selected as the SII required in the dark zone.

From a design perspective, the SII-based approach to selecting the SNR is attractive as it eliminates the need to conduct listening tests, but the limitations of the method must be taken into account. Table I shows the SNR required to achieve privacy, according to either the measured SRT values, or conditions where the SII = 0.05. Although this SII value represents the average SRT across all tested array geometries and conditions, as described above, there is a 5 dB difference in the required SNR between the two methods in the case of the wide array. This can be attributed to the design of the SII algorithm, which aggregates SNRs from several frequency bands into a single number rating using a weighted average. If an array aliases within the speech frequency range, and this results in a reduced contrast, for example as shown in Figure 3 at 3 kHz for the wide array, then the SNR in this band is greater than in other bands. Consequently, some speech sounds can be more easily understood than others, increasing the intelligibility over that predicted by the weighted average SNR[47]. It is recommended therefore that array designs that exhibit sharp reductions in acoustic contrast within the speech frequency range, due to spatial aliasing or other effects, be avoided as additional masking is required to compensate, and the true intelligibility is harder to predict using standard metrics.

Condition	SNR_d (dB)	
	$\mathrm{SII}_d = 0.05$	SRT
W_S	-16	-21
W_A	-27	-32
W_∞	-12	-17
N_S	-13	-15
N_A	-12	-12
N_∞	-12	-10

TABLE I. Comparison between dark zone signal-to-noise ratios, SNR_d, required for privacy when estimated using SII simulations and experimental SRTs.

B. Identification of Feasible Masking Signals

The listening test results described above show the SNR that is required in the dark zone to achieve privacy at three low-pass filter settings for each array configuration. However, this SNR can also be calculated using

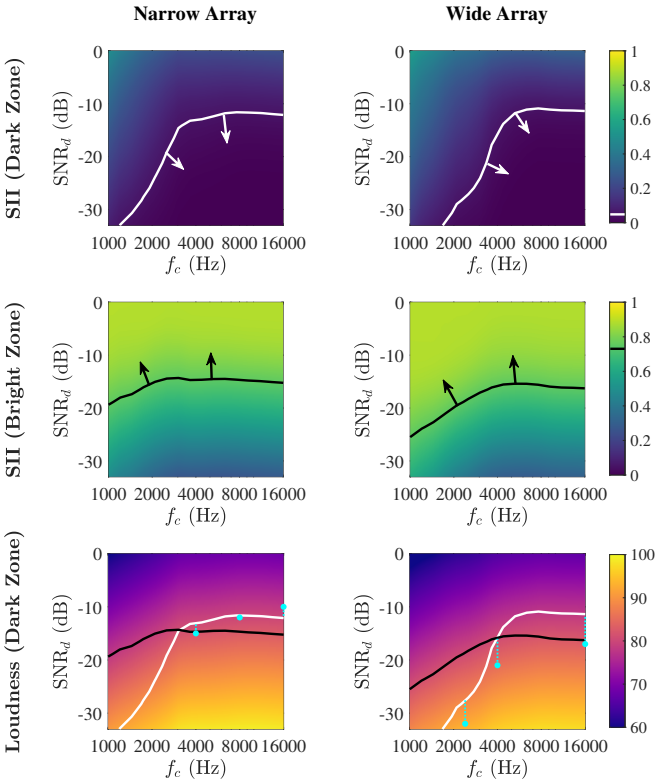

FIG. 7. [Colour Online] Surface plots of SII and Loudness with variation in dark zone signal-to-noise ratio and masking signal cut-off frequency. Upper Row: SII in dark zone. Middle Row: SII in bright zone. Lower Row: Loudness of masker in dark zone. White line: $\mathrm{SII}_d = 0.05$, Black line: $\mathrm{SII}_b = 0.75$. Arrows indicate regions of the parameter space where intelligibility constraints are met.

SII=0.05 as a proxy for privacy, using auralisations of the speech and noise emitted from the array, recorded in each zone. Signal auralisations are not strictly necessary for this purpose, as the SII algorithm only uses the spectra of the speech and noise signals to form its intelligibility estimate. These spectra can be synthesised by combining the frequency response of the array with the predicted acoustic contrast and standard speech spectra. The adjustments to the speech-shaped maskers presented in the test can be described using two parameters: the masker level and the cut-off frequency of the low-pass filter. By generating auralisations across a range of values of these two parameters, and analysing the SII in the bright and dark zones at each data point, contours of bright and dark zone SII can be generated. These are shown for the narrow and wide array in Figure 7.

Each point on each surface represents the outcome from a single simulation, where the masker has been low-pass filtered with some cut-off frequency, f_c, and has had its gain adjusted to produce a given signal-to-noise ratio in the dark zone, SNR_d. This representation of the masker level is chosen to facilitate comparison with the listening test results presented in Section VI. The upper row of plots shows the SII in the dark zone, for the narrow array on the left and the wide array on the right. The middle row shows the corresponding SII in the bright zone. The intelligibility constraints $\mathrm{SII}_d < 0.05$ and $\mathrm{SII}_b > 0.75$ are visualised as contour lines of equal intelligibility through the parameter space in the upper and middle plots respectively. All points in the parameter space below the white contour at $\mathrm{SII} = 0.05$ represent situations with sufficiently low intelligibility in the dark zone to claim privacy. Likewise, all points above the bright zone constraint contour, shown by the black line at $\mathrm{SII}_b = 0.75$, exceed the ANSI guideline for "good" speech reproduction in the bright zone[35].

From the results presented in Figure 7, it can be seen that in order to provide speech privacy with a masker whose filter cut-off frequency, f_c, is low, the SNR must be reduced significantly as the masker and speech signal do not overlap sufficiently in frequency for the masker to be effective. As f_c is increased, the speech and masker spectra become more similar, so the masker gain can be reduced (i.e. the SNR increased) whilst maintaining the same predicted intelligibility level. Above 5 kHz, the contours become approximately constant with an increase in f_c. This is a feature of the SII metric, which assigns each critical band an importance value. Above 4.8 kHz, the relative importance of each critical band to speech intelligibility sharply decreases, so changes in the difference between the speech and noise spectra have little impact on the final value of the SII.

The intelligibility contours from the upper four plots in Figure 7 are transferred onto the lower row of plots to provide an enclosed *feasible region* in which both intelligibility constraints are met. The light blue points in the lower panels of Figure 7 represent the experimental SRT values from Table I. The dashed lines indicate the difference between the SNR required to achieve the measured SRT and SII=0.05 at the three filter cut-off frequencies. From this lower set of plots it can be seen that the optimal masking signal parametrisation within the feasible region can be decided by considering the perceptual attributes of the dark zone sound field, a decision which is guided by preference test results and subjective metrics. This will be discussed further in the following section.

It is important to highlight that the contours presented here are specific to the position of the zones, the loudspeaker array used and the surrounding room acoustics. However, the discussion relating to the relative positions of the contour lines and the behaviour with different spatial aliasing characteristics should hold for a range of multi-zone problems, and therefore provide general insight into the privacy control design problem. Certain situations can cause there to be no intersection between the regions where $\mathrm{SII}_b > 0.75$ and $\mathrm{SII}_d < 0.05$, such as when the size of the array (in terms of the array length, D, or the number of elements, L) prevents sufficient acoustic contrast from being provided, or room reverberation is too high. When this is the case, either the constraints on the bright and/or dark zone intelligibility must be made less onerous or the system must be redesigned.

Condition A	vs.		Condition B
W_S	97.4%	2.6%	W_A
W_S	11.8%	88.2%	W_∞
W_A	0.4%	99.6%	W_∞
N_S	22.3%	77.7%	N_A
N_S	22.7%	77.3%	N_∞
N_A	50.5%	49.5%	N_∞

TABLE II. Percentage likelihood of one condition being preferred over another, using data gathered from the preference tests.

VII. PREFERENCE TEST RESULTS AND DISCUSSION

The paired preference tests, as described in Section V B have been analysed using statistical tools by Perez-Ortiz and Mantiuk[48]. These tools analyse patterns of preference across participants to enable outliers to be removed, indicating potential misunderstandings of the task. The raw preference results are converted to percentage likelihoods of one condition being chosen over another, allowing both an ordering of conditions from best to worst and an analysis of how significant those preferences are. A preference of greater than 75% for one condition over another is deemed to be significant[48]. The results for each of the comparisons are presented in Table II.

The preference test results for the wide array were conclusive, with participants demonstrating a clear preference in all three paired comparisons. The least preferred condition was W_A, where the low-pass filter is set at 2.4 kHz to prevent aliasing. Figure 6 shows that the median SRT at this condition is -32 dB, at least 10 dB lower than the other two conditions in the test. The corresponding increase in the masker level was clearly perceivable, and was disliked by participants. In the comparison between W_S and W_∞, a significant proportion of participants selected W_∞, the case where the masker has broader bandwidth and a lower overall level. Preferences were not as well-defined in the case of the narrow array, suggesting that the signals in each condition were perceptually more similar than those with the wide array. Applying the low-pass filter to reduce sharpness (N_S) was disliked when compared with both the unfiltered condition N_∞ and against N_A, where the low-pass filter cut-off was set to 8 kHz to prevent spatial aliasing. This can again be related to the increased masker level required when the cut-off frequency of the low-pass filter is reduced to 4 kHz. No significant preference was shown between the N_∞ and N_A conditions when preferences are aggregated across participants, but more than half (11 of 21) listeners consistently chose their preferred condition across all four repeats of the N_A vs N_∞ test. Only three participants selected each condition twice (equivalent to chance), indicating that for most listeners, it was possible to distinguish the samples and make a repeatable preference judgement.

After the completion of each preference test, participants were asked to comment on any features of the noise samples that they were listening for. Despite the apparent correlation between masking signal level and preference judgement, only 8 out of the 21 participants mentioned loudness or quietness in their descriptions. 17 participants distinguished between sounds by referring to their spectrum, using words such as “high/low pitched”, “sharpness” or “harshness”. 10 participants used words referring to naturalness or artificiality, e.g. “smooth”, “natural”, “sounds like a jet engine / waterfall / the London Underground”. When such a preference was expressed, participants unanimously preferred sounds they deemed to be “natural” over those which were “artificial”. When correlated with the participants’ individual choices in the preference test, sounds with broader bandwidth were deemed to sound more “natural”. Six participants commented after the narrow array preference test that although masking signals with wider bandwidth were preferred in general, having too much high frequency content was detrimental. Of these six, five preferred the condition N_A over N_∞, backing up their comments with their preference decisions. Although participants only listened to filtered random noise samples, a surprising breadth of semantic descriptions were attached to these sounds. This encourages future experimentation on the acceptability of different types of masking signals in different contexts.

A. Comparison with Subjective Metrics

Subjective metrics are designed to quantify perceptual features from signals, and are useful alternatives to costly, complex jury testing. Table III shows the values of the Psychoacoustic Annoyance, Loudness, Roughness and Sharpness metrics when applied to the stimuli presented in the preference test; highlighted cells indicate where the metric predicts the order of preference from the test results. The complete psychoacoustic annoyance metric fails to predict the order of preference in the case of the narrow array, and the standard deviation of the results is also large compared to that of the other metrics. This is to be expected, as the uncertainty of the combined metric will include the uncertainties of the sub-metrics. Although the metrics themselves are deterministic, the uncertainty stems from the fine structure of the randomly generated speech and noise signals. Sharpness, which was hypothesised during test development to be an undesirable attribute, was instead found to be inversely related to the preference results; signals with higher sharpness values were preferred on average. However, attention must be paid to the absolute sharpness value. The formulation of the sharpness model[19] states that this attribute only affects psychoacoustic annoyance when it exceeds 1.75 acum (Eq. 2); only two of the tested values exceed this threshold as speech-shaped noise contains relatively little energy above 3 kHz, where

sharpness begins to be perceived. Comments from participants who selected the case N_A over N_∞ suggest that the increased sharpness of the N_∞ condition was undesirable.

The loudness and roughness metrics perform better overall, with the roughness metric correctly predicting the preference order of all conditions and the loudness metric correctly identifying one of the most preferred options in the case of the narrow array. Cursory inspection of Table III suggests that the roughness metric is the superior predictor, but the absolute value and just noticeable difference (JND) for roughness perception must also be taken into account. The roughness of unmodulated noise is caused by random amplitude fluctuations and is therefore dependent on its bandwidth[49]. For example, peak roughness between 0.2 and 0.3 asper occurs for noise with a bandwidth of 100 Hz, decreasing thereafter to around 0.05 asper at full audio bandwidth. For amplitude modulated tones, the threshold of roughness perception is 0.07 asper, and the just noticeable difference limen $\Delta R/R$ is 17%. Therefore, while the roughness values of the stimuli can be distinguished from one another, the absolute roughness level of all the tested stimuli is already very low. Roughness is difficult to explicitly control without also affecting other perceptual features, as no explicit amplitude modulation is included in the masker. Furthermore, the roughness of a given signal increases slightly with signal level[49], and is thus nonlinear. This is exemplified in the case of the wide array where, although the roughness metric correctly predicts the order of preference, this effect is due to the large level difference between stimuli. The recommendation is, therefore, that a masking signal should be determined based primarily on minimising loudness, with attention also being paid towards minimising the residual roughness of the masker. This further motivates investigation into the context-dependence of masker preference, as the preference for "naturalness" expressed by participants appears to be well-modelled by the roughness metric when applied to stationary random noise. An investigation using a range of natural masking sounds may also shed light on the advantages and disadvantages of masker fluctuation, which is negligible for the stimuli tested here.

B. Selection of Preferred Masker

Analysis of the preference test results shows that loudness and roughness should be minimised when selecting a masking signal. The surface plots in the lower two panels of Figure 7 show the variability of loudness with respect to the dark zone SNR and the masker cut-off frequency. System designers can be guided by generating equivalent contour and surface plots for candidate array designs. Within the feasible region enclosed by the upper and lower intelligibility constraints, the point with the lowest loudness should be selected. This point usually lies on the dark zone constraint boundary (white contour), as masking signals that just satisfy this constraint have just enough energy to provide sufficient masking.

Condition, pref. order	Annoyance	Loudness	Roughness	Sharpness
W_∞, 1	80.0±0.5	77.7±0.2	0.04±0.003	1.61±0.01
W_S, 2	84.8±0.6	81.3±0.2	0.06±0.003	1.21±0.08
W_A, 3	91.4±1.1	85.7±0.3	0.08±0.006	1.05±0.08
N_∞, =1	84.8±2.0	79.3±0.1	0.04±0.003	1.88±0.08
N_A, =1	83.2±1.8	79.6±0.1	0.04±0.003	1.80±0.06
N_S, 2	82.8±0.7	79.7±0.2	0.05±0.003	1.58±0.07

TABLE III. Metric values of stimuli presented to participants in the paired-preference test. Uncertainties are $\pm 1\sigma$. Highlighted cells indicate where the metric correctly predicts the order of preference within each array width.

In areas where the loudness contours are approximately constant under adjustment of one parameter, this may indicate that the corresponding signals are physically similar, for instance, low-pass filtered speech shaped noise with cut-off frequencies between 8 and 16 kHz. Analysis of the predicted roughness at positions with similar loudness can provide fine-tuning to the selected masker, though this requires a full auralisation of the signal in the dark zone, rather than simply an estimate of the spectrum of the masker reproduced in the dark zone, which is sufficient for the estimation of loudness.

VIII. SUMMARY OF DESIGN METHOD

The overall structure of the proposed design process is described in block diagram form in Figure 8. The method requires an estimate of the transfer responses from a candidate array design to the designated bright and dark zones. These transfer responses, whether derived from measurements or simulations, enable the production of sound zoning filters, which can be used to create auralisations of speech and masker signals in each zone. Speech intelligibility is evaluated in both zones using the Speech Intelligibility Index (SII) metric[35], to determine the range of reproduction levels that satisfy the SII constraints of SII$<$ 0.05 in the dark zone and SII$>$ 0.75 in the bright zone. If these constraints can be met simultaneously, the masker with the lowest loudness,

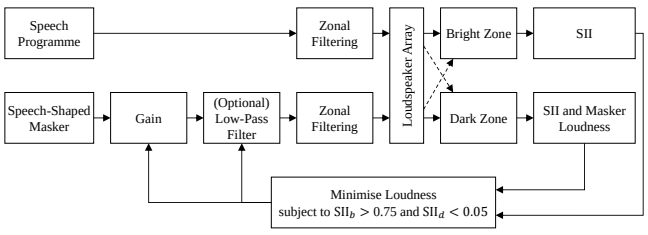

FIG. 8. Block diagram of the personal audio system design method.

evaluated in the dark zone can be selected. Otherwise, the array geometry or the positions of the zones are potentially unsuitable for effective private sound zone reproduction, and alternatives that offer greater acoustic contrast should be considered.

A further consideration regarding the overall acceptability of a designed system, is that the dominant output from the system should be speech, rather than noise. In other words, the level of the masking signal in the dark zone should be kept below that of the speech programme in the bright zone. Systems where this is not the case are likely to be condemned as sources of noise pollution, as this may impede conversation between occupants of the dark zone. This condition can be checked by comparing the acoustic contrast provided by the array across the speech bandwidth against the SNR required for privacy. For the two presented configurations, the acoustic contrast is between 15 and 20 dB across the speech frequency range, and the required SNR_d to ensure privacy is -12 dB for both configurations, using unfiltered speech-shaped maskers. This means that the required level of the masker in the dark zone is lower than that of the speech in the bright zone.

IX. CONCLUSIONS

The creation of private sound zones can be attained by radiating a masking signal into the dark zone of a conventional personal audio system. Successful systems must adhere to speech intelligibility constraints in each zone, whilst also considering the potential for the masking signal to become annoying to listeners in the dark zone. Listening test results, objective intelligibility predictions and subjective metrics have been combined to form guidance on how to design such a system to achieve these targets.

The horizontal spacing of source array elements can lead to spatial aliasing of the masker within the speech frequency range. One strategy to prevent aliased side-lobes from compromising the bright zone sound field, proposed by Donley et. al.[17], is to apply a low-pass filter to the masker. However, this impedes the effectiveness of the masker, requiring higher masker levels to achieve the same intelligibility contrast. Listening test results indicate that in system geometries where spatial aliasing in the speech bandwidth is unavoidable, it is more acceptable to use a broadband speech-shaped masker that can be reproduced at a lower level, compared to eliminating the spatially aliased component with a low-pass filter. It is recommended to perform in-situ directivity measurements above the aliasing frequency to determine whether spatially aliased side-lobes, or any associated specular reflections, impinge on the bright zone.

The symmetric zonal layout and stationary, random noise maskers used in this work were intended to mitigate contextual biases associated with different listening scenarios. Despite this, listeners likened the masking sounds presented in the preference test to a wide range of natural and unnatural environmental sounds. Correlation of these preference results with perceptual metrics suggests that the reduction of subjective roughness is important. This prompts further investigation into the acceptability of different types of masking signals, particularly in conjunction with a study on how this is affected by context.

ACKNOWLEDGMENTS

This work was supported by an EPSRC Doctoral Training Centre grant (EP/L015382/1). Loudness and Sharpness models used the GENESIS Loudness Toolbox, and the implementation of the Roughness model was by Matt Flax at the Psysound3 project. Code for these functions is available at the MATLAB File Exchange[50].

[1]W. F. Druyvesteyn and J. Garas, "Personal Sound," Journal of the Audio Engineering Society **45**(9), 685–701 (1997) http://www.aes.org/e-lib/inst/browse.cfm?elib=7843.

[2]M. F. Simon Galvez, S. J. Elliott, and J. Cheer, "Personal Audio Loudspeaker Array as a Complementary TV Sound System for the Hard of Hearing," IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences **E97.A**(9), 1824–1831 (2014) doi: https://doi.org/10.1587/transfun.E97.A.1824.

[3]J.-H. Chang, C.-H. Lee, J.-Y. Park, and Y.-H. Kim, "A realization of sound focused personal audio system using acoustic contrast control.," The Journal of the Acoustical Society of America **125**(4), 2091–2097 (2009) doi: 10.1121/1.3082114.

[4]K. Baykaner, C. Hummersone, R. Mason, and S. Bech, "The acceptability of speech with interfering radio program material," in *Audio Engineering Society Convention 136* (2014), http://www.aes.org/e-lib/browse.cfm?elib=17167.

[5]J. Cheer, S. J. Elliott, and M. F. S. Gálvez, "Design and implementation of a car cabin personal audio system," Journal of the Audio Engineering Society **61**(6), 412–424 (2013).

[6]S. J. Elliott and M. Jones, "An active headrest for personal audio," The Journal of the Acoustical Society of America **119**(5), 2702 (2006) doi: 10.1121/1.2188814.

[7]S. J. Elliott, J. Cheer, H. Murfet, and K. R. Holland, "Minimally radiating sources for personal audio," The Journal of the Acoustical Society of America **128**(4), 1721–1728 (2010) doi: 10.1121/1.3479758.

[8]J.-W. Choi and Y.-H. Kim, "Generation of an acoustically bright zone with an illuminated region using multiple sources," The Journal of the Acoustical Society of America **111**(4), 1695–1700 (2002) doi: 10.1121/1.1456926.

[9]F. Jacobsen, M. Olsen, M. Møller, and F. T. Agerkvist, "A Comparison of Two Strategies for Generating Sound Zones in a Room," in *Proc. 18th International Congress on Sound and Vibration*, Rio De Janiero (2011), https://orbit.dtu.dk/files/5677256/ICSV18FJ[1].pdf.

[10]K. Baykaner, P. Coleman, R. Mason, P. J. Jackson, J. Francombe, M. Olik, and S. Bech, "The relationship between target quality and interference in sound zones," AES: Journal of the Audio Engineering Society **63**(1-2), 78–89 (2015) doi: 10.17743/jaes.2015.0007.

[11]J. Francombe, P. Coleman, M. Olik, K. Baykaner, P. J. B. Jackson, R. Mason, S. Bech, J. A. Pedersen, M. Dewhirst, and J. A. Pederson, "Perceptually Optimized Loudspeaker Selection for the Creation of Personal Sound Zones," 52nd Audio Engineering Society Conference (2013) http://www.aes.org/e-lib/browse.cfm?elib=16907.

[12]P. Coleman, P. J. B. Jackson, M. Olik, and J. Abildgaard Pedersen, "Personal audio with a planar bright zone," The Journal of the Acoustical Society of America **136**(4), 1725–1735 (2014) doi: 10.1121/1.4893909.

[13]F. Olivieri, F. M. Fazi, S. Fontana, D. Menzies, and P. A. Nelson, "Generation of Private Sound with a Circular Loudspeaker Array and the Weighted Pressure Matching Method," IEEE/ACM Transactions on Audio Speech and Language Processing **25**(8), 1579–1591 (2017) doi: 10.1109/TASLP.2017.2700945.

[14]T. Lee, J. K. Nielsen, and M. G. Christensen, "Towards Perceptually Optimized Sound Zones: A Proof-of-Concept study," in *IEEE International Conference on Acoustics, Speech and Signal Processing*, IEEE, Brighton, UK (2019), pp. 136–140, doi: https://doi.org/10.1109/ICASSP.2019.8682902.

[15]J. Donley, C. Ritz, and W. B. Kleijn, "Improving Speech Privacy in Personal Sound Zones," in *IEEE International Conference on Acoustics, Speech and Signal Processing* (2016), pp. 311–315.

[16]K. Chwioko, D. Nourzad, and X. Vinamata, "Apparatus and Method for Privacy Enhancement," (2018), https://patents.google.com/patent/WO2018046185A1.

[17]J. Donley, C. H. Ritz, and W. B. Kleijn, "Multizone Soundfield Reproduction With Privacy and Quality Based Speech Masking Filters," IEEE/ACM Transactions on Audio Speech and Language Processing **26**(4), 1–15 (2018) doi: 10.1109/TASLP.2018.2798804.

[18]D. Wallace and J. Cheer, "Optimisation of Personal Audio Systems for Intelligibility Contrast," in *Proc. 144th Audio Engineering Society Convention*, Milan, Italy (2018).

[19]H. Fastl and E. Zwicker, *Psychoacoustics: Facts and models*, 3rd ed. (Springer, 2007), p. 328.

[20]D. Wallace and J. Cheer, "Combining Artificial and Natural Background Noise in Personal Audio Systems," in *Proc. 10th IEEE Sensor Array and Multichannel Signal Processing Workshop*, Sheffield, UK (2018).

[21]C. House, S. Dennison, D. G. Morgan, N. Rushton, G. V. White, J. Cheer, and S. Elliott, "Personal Spatial Audio in Cars Development of a loudspeaker array for multi-listener transaural reproduction in a vehicle," in *Proceedings of the Institute of Acoustics* (2017), Vol. 39. pt. 2.

[22]H. Kuttruff, *Room Acoustics*, 3rd ed. (Elsevier, 1991), p. 123.

[23]M. Olik, J. Francombe, P. Coleman, P. J. B. Jackson, M. Olsen, M. Møller, R. Mason, and S. Bech, "A comparative performance study of sound zoning methods in a reflective environment," in *52nd Audio Engineering Society Conference* (2013), http://www.aes.org/e-lib/browse.cfm?elib=16914.

[24]M. A. Akeroyd, J. Chambers, D. Bullock, A. R. Palmer, A. Q. Summerfield, P. A. Nelson, and S. Gatehouse, "The binaural performance of a cross-talk cancellation system with matched or mismatched setup and playback acoustics," The Journal of the Acoustical Society of America **121**(2), 1056–1069 (2007) doi: 10.1121/1.2404625.

[25]M. Olik, P. J. Jackson, and P. Coleman, "Influence of low-order room reflections on sound zone system performance," in *Proceedings of Meetings on Acoustics* (2013), Vol. 19, doi: 10.1121/1.4800873.

[26]N. R. French and J. C. Steinberg, "Factors Governing the Intelligibility of Speech Sounds," Journal of the Acoustical Society of America **19**(1), 90–119 (1947) doi: 10.1121/1.1916407.

[27]J. M. Festen and R. Plomp, "Effects of fluctuating noise and interfering speech on the speech-reception threshold for impaired and normal hearing," Journal of the Acoustical Society of America **88**(4), 1725–1736 (1990) doi: 10.1121/1.400247.

[28]A. W. Bronkhorst, "The cocktail-party problem revisited: early processing and selection of multi-talker speech," Attention, Perception, and Psychophysics **77**(5), 1465–1487 (2015) doi: 10.3758/s13414-015-0882-9.

[29]J. S. Bradley, "The acoustical design of conventional open plan offices," Canadian Acoustics - Acoustique Canadienne **31**(2), 23–31 (2003).

[30]W. J. Cavanaugh, W. R. Farrell, P. W. Hirtle, and B. G. Watters, "Speech Privacy in Buildings," The Journal of the Acoustical Society of America **34**(4), 475–492 (1962) doi: 10.1121/1.1918154.

[31]ASTM International, "ASTM E1130-16: Standard Test Method for Objective Measurement of Speech Privacy in Open Plan Spaces Using Articulation Index," (2016).

[32]K. S. Rhebergen, J. Lyzenga, W. A. Dreschler, and J. M. Festen, "Modeling speech intelligibility in quiet and noise in listeners with normal and impaired hearing," The Journal of the Acoustical Society of America **127**(3), 1570–1583 (2010) doi: 10.1121/1.3291000.

[33]A. C. C. Warnock, "Acoustical privacy in the landscaped office," Journal Of The Acoustical Society Of America **53**(6), 1535–1543 (1973) doi: 10.1121/1.1913498.

[34]B. N. Gover and J. S. Bradley, "Measures for assessing architectural speech security (privacy) of closed offices and meeting rooms," The Journal of the Acoustical Society of America **116**(6), 3480–3490 (2004) doi: 10.1121/1.1810300.

[35]ANSI, "ANSI/ASA S3.5-1997 (R2017) Methods for Calculation of the Speech Intelligibility Index," (1997).

[36]U. Widmann, "A Psychoacoustic Annoyance Concept for Application in Sound Quality," in *Proc. Noise-Con 1997* (1997).

[37]E. Tristán-Hernández, I. P. García, J. M. L. Navarro, I. Campos-Cantón, and E. S. Kolosovas-Machuca, "Evaluation of psychoacoustic annoyance and perception of noise annoyance inside university facilities," International Journal of Acoustics and Vibrations **23**(1), 3–8 (2018) doi: 10.20855/ijav.2018.23.11059.

[38]M. Schneider and C. Feldmann, "Psychoacoustic evaluation of fan noise," FAN 2015 – International Conference on Fan Noise, Technology and Numerical Methods (2015).

[39]K. Fujii, J. Atagi, and Y. Ando, "Temporal and spatial factors of traffic noise and its annoyance," Journal of Temporal Design in Architecture and the Environment **2**(1), 33–41 (2002) http://www.jtdweb.org/journal/2002/005_fujii.pdf.

[40]S. Rosen, P. Souza, C. Ekelund, and A. A. Majeed, "Listening to speech in a background of other talkers: Effects of talker number and noise vocoding," The Journal of the Acoustical Society of America **133**(4), 2431–2443 (2013) doi: 10.1121/1.4794379.

[41]S. J. Hall, "The Development of a New English Sentence in Noise Test and an English Number Recognition Test," (2006), MSc Thesis, University of Southampton.

[42]D. R. Hewitt, "Evaluation of an English Speech-in-Noise Audiometry Test," (2008), MSc Thesis, University of Southampton.

[43]B. Hagerman, "Sentences for Testing Speech Intelligibility in Noise," Scandinavian Audiology **11**(2), 79–87 (1982) doi: 10.3109/01050398209076203.

[44]K. Wagener, V. Kühnel, and B. Kollmeier, "Development and Evaluation of a German Sentence Test," Zeitschrift für Audiologie **31**(1), 1–32 (1999).

[45]B. Kollmeier, A. Warzybok, S. Hochmuth, M. A. Zokoll, V. Uslar, T. Brand, and K. C. Wagener, "The multilingual matrix test: Principles, applications, and comparison across languages: A review," International Journal of Audiology **54**, 3–16 (2015) doi: 10.3109/14992027.2015.1020971.

[46]T. Brand and B. Kollmeier, "Efficient adaptive procedures for threshold and concurrent slope estimates for psychophysics and speech intelligibility tests," The Journal of the Acoustical Society of America **111**(6), 2801–2810 (2002) doi: 10.1121/1.1479152.

[47]T. Leclère, D. Théry, M. Lavandier, and J. F. Culling, "Speech intelligibility for target and masker with different spectra," in *Physiology, Psychoacoustics and Cognition in Normal and Impaired Hearing*, Springer International Publishing, Cham (2016), pp. 257–266.

[48]M. Perez-Ortiz and R. K. Mantiuk, "A practical guide and software for analysing pairwise comparison experiments," (2017), http://arxiv.org/abs/1712.03686.

[49]P. Daniel and R. Weber, "Psychoacoustical Roughness: Implementation of an Optimized Model," Acta Acustica united with Acustica **83**(1), 113–123 (1997).

[50]S. Bleeck, "matlab_real_time_sound," (2020), https://uk.mathworks.com/matlabcentral/fileexchange/71950-matlab{_}real{_}time{_}sound.