Audio Engineering Society

Convention Paper

Presented at the 144th Convention
2018 May 23 – 26, Milan, Italy

Optimisation of Personal Audio Systems for Intelligibility Contrast

Daniel Wallace[1] and Jordan Cheer[1]

[1] *University of Southampton*

Correspondence should be addressed to Daniel Wallace (djw1g12@soton.ac.uk)

ABSTRACT

Personal audio systems are designed to deliver spatially separated regions of audio to individual listeners. This paper demonstrates a method of personal audio system design which provides a level of contrast in the perceived speech intelligibility between bright and dark audio zones. Limitations in array directivity which would lead to a loss of privacy are overcome by reproducing a synthetic masking signal in the dark zone. This signal is optimised to provide effective masking whilst remaining subjectively pleasant to listeners. Results of this optimisation from a simulated personal audio system are presented.

INTRODUCTION

The design of personal audio systems, which direct sound to a target listener, must take into account both physical acoustics and psychoacoustics to achieve the highest level of perceived performance. Such systems find utility in shared office spaces and museum exhibits [1], television systems [2], headrest-mounted loudspeaker systems [3], in-car entertainment [4] and mobile devices [5]. This paper shows how a personal audio system may be designed to provide two zones of sound, designated as acoustically *bright* and *dark*. However, in this work, the contrast is defined as the inter-zone difference in intelligibility of a speech signal, rather than as a measure of the difference in energy between each zone.

In order to reduce the intelligibility of the programme signal by unintended listeners, the proposed personal audio system radiates a masking signal into the dark zone. Two acoustic contrast control processes [6] are combined to produce this result. The first aims to maximise the level of the speech programme in the bright zone whilst minimising its radiation into the dark zone, and the second maximises the level of the masking signal in the dark zone, whilst minimising its intrusive effect on the programme in the bright zone.

For a system to be successful, it is not sufficient for the masking signal to simply provide privacy between zones, as high masker levels may result in unnecessary noise pollution in the vicinity of the system. An optimisation procedure is therefore necessary to design the signal, with the objective of simultaneously providing adequate intelligibility difference between the zones and minimising the potential for annoyance in the dark zone.

Throughout this paper, results from a simulated loudspeaker array are presented. The system is simulated using an array of point monopole sources, with point omnidirectional receivers demarcating the two audio

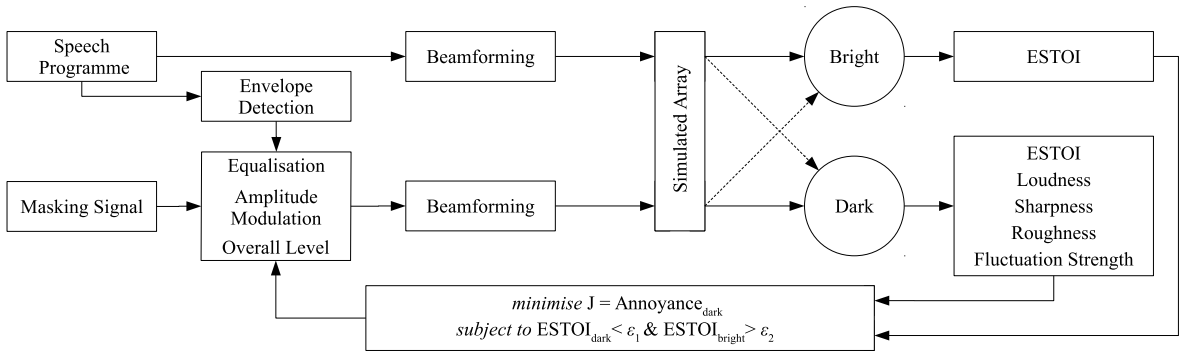

Fig. 1: Block Diagram of the proposed personal audio system. Speech is focussed into the bright zone using optimal acoustic contrast control. The masking signal is modified in terms of spectrum and overall sound pressure level, and envelope detection is used to activate and deactivate the masker when speech is present. The parameters which control these modifications are updated by a constrained optimisation loop which minimises the estimated annoyance in the dark zone whilst maintaining a minimum level of intelligibility (ESTOI) contrast set by ε_1 and ε_2.

zones. Information about the performance of the system is ascertained through the use of subjective metrics, which undertake to provide a mapping between the measurable, objective parameters of a signal and the expected subjective response from a population. The optimisation relies on the Extended Short-Time Objective Intelligibility (ESTOI) [7] and Psychoacoustic Annoyance [8] metrics. The latter is constructed from metrics for loudness, sharpness, roughness and fluctuation strength. Figure 1 shows a block diagram of the optimisation loop employed in the design of the personal audio system demonstrated in this paper.

Firstly, attention is paid to the design and performance of the loudspeaker array which underpins the personal audio system. This is followed by a discussion of the subjective metrics which are used to evaluate the quality of the system. The process of optimising the masking signal is then discussed, followed by associated results and conclusions.

ARRAY DESIGN

The psychoacoustic concern of privacy between audio zones can be attributed to physical limitations in loudspeaker array design. The level of acoustic contrast achievable using a certain system geometry fundamentally limits the amount of control a system designer has over the signals in the bright and dark zones. This is evidenced by considering a mathematical derivation of the acoustic contrast control process, such as that provided in Section II B. of [9].

Following the notation in [9], the array consists of L drivers with weights $\mathbf{u}$. In theory, arbitrary levels of acoustic contrast can be achieved by increasing L and the array effort, which is proportional to the total weight power $\mathbf{u}^H\mathbf{u}$. A Tikhonov regularisation parameter is included to constrain array effort and improve the numerical stability of the simulation. This also results in simulated levels of acoustic contrast corresponding more closely to that measured from physical arrays with power handling limits, imperfect matching between drivers, and sensitivity to changes in the environment which are not explicitly simulated.

SYSTEM GEOMETRY

The simulated array used in this paper has $L = 4$ elements. Bright and dark zones are symmetric with respect to the array, and are each formed by nine point receivers. Figure 2 shows the positions of the zones with respect to the array elements. The symmetry of the geometry is significant as, in general, the performance plots in Figures 4 and 5 would exhibit differences depending on whether the solution is being calculated for the beamforming process which targets programme material into the bright zone or the masking signal into the dark zone; however, here the solutions are identical due to the symmetrical geometry.

The array regularisation parameter is initially set to vary with frequency, increasing exponentially from a value of 10^{-4} at low frequency to 10^{2} at high frequency, indicated with a dashed line in Figure 3. This is to ensure robustness at high frequencies where practical variations between loudspeaker elements will be significant, and to limit drive levels at low frequency. At frequencies where an array effort greater than 6 dB is demanded by the acoustic contrast control process, the regularisation parameter is increased from its initial setting. This has the effect of reducing acoustic contrast and flattening the low frequency response of the array, which otherwise requires excessive drive levels.

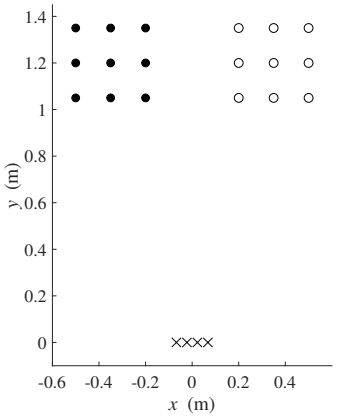

Fig. 2: Personal audio system geometry. Array Elements: ×, Dark zone microphones: •, Bright zone microphones: ○. The spacing of array elements is 0.045 metres.

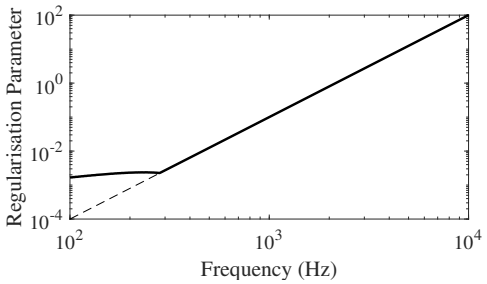

Fig. 3: Frequency dependence of the regularisation parameter. Where array effort would otherwise exceed 6 dB, i.e. below 300 Hz, the regularisation parameter is increased from its initial parametrisation, indicated with a dashed line.

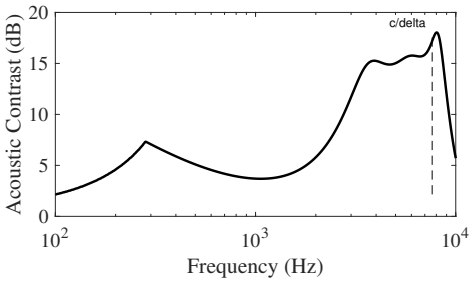

Fig. 4: Acoustic Contrast between bright and dark zones. Contrast is reduced at low frequencies due to the limited array effort. c/delta is the aliasing limit for the array; the ratio of sound speed to the spacing of array elements.

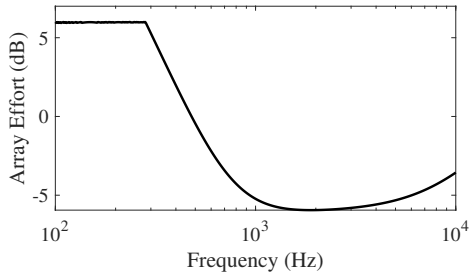

Fig. 5: Array Effort required to provide acoustic contrast. At low frequencies, array effort increases due to cancellation between array driver elements. Increased regularisation below 300 Hz limits array effort at 6 dB.

SUBJECTIVE METRICS

In order to quantify the subjective performance of the personal audio system, a number of metrics are combined to give an overall subjective impression. The aim of any subjective metric is to provide a mapping between measurable, objective parameters of a signal, the *stimulus*, and the expected subjective response, the *sensation* experienced by an average person. Subjective metrics can be loosely ranked or categorised based on the strength of the relation between stimulus and sensation. For example, the subjective experience of loudness is primarily related to intensity, with further spectral and temporal effects. Following the well known work of Zwicker and Fastl [8], the high level impression of annoyance is broken down into four elementary metrics; loudness, sharpness, roughness and fluctua-

tion strength, each of which have various dependencies on properties of the signal. A metric that corresponds to the intelligibility of speech signals in the bright and dark zones is used to assess the degradation and privacy of the target speech material.

INTELLIGIBILITY

A number of metrics that can estimate the intelligibility of a speech signal exist, differing in implementation and intended application. Some algorithms such as PESQ-FR [10], SII [11] and STOI [12] compare a reference signal with a degraded signal. Conversely, the PESQ-NR [13] and Speech Transmission Index [14] algorithms are *single-ended* measurements, comparing the statistics of the degraded signal with assumed properties such as the spectrum and modulation pattern of speech. For the application discussed in this paper, algorithms of the former class can be used as the personal audio system can retain a reference copy of the speech programme material sent to the array.

For maximum flexibility, the designed personal audio system should be capable of assessing the intelligibility of the speech programme in the bright and dark zones after degradation caused by an arbitrary masking signal. Consequently, an appropriate intelligibility metric must be consistent under many different forms of additive noise, as well as spectral shaping of the signals by the array. Many intelligibility algorithms make use of the global statistics of a signal, which results in poor estimation of speech intelligibility when additive noise is time-varying [12]. The Extended Short-Time Objective Intelligibility (ESTOI) algorithm [7] is designed to overcome this limitation by dividing the signals into 384 ms segments, a value chosen to ensure the algorithm is sensitive to important temporal modulation above 2.6 Hz. The algorithm forms an intelligibility rating by calculating correlation coefficients between short time spectrograms of the original and degraded signal across frequency. This intermediate intelligibility index for each 384 ms frame is then averaged over time to produce a scalar output for a given pair of input signals.

Figure 6 shows a block diagram of the algorithm. As ESTOI is calculated from sequential frames of audio, it can potentially be adapted to be used for real-time optimisation of the masking signal.

ANNOYANCE

Typical handling of acoustic annoyance for environmental noise is limited to controlling sound level; European Union guidelines [15] for the control of environmental noise cite the annoyance caused by noise as one motivation for recommending an upper bound on the time averaged A-weighted sound pressure level during night-time hours. The use of such crude measures to capture annoyance is often justified by the additional time, expense and computation to process time histories required by more complex metrics [16]. A well known and widely used annoyance metric was proposed by Zwicker and Fastl [8]. Their *Psychoacoustic Annoyance* requires the computation of Loudness, Sharpness, Roughness and Fluctuation Strength in the following combination:

$$PA = N_5 \left(1 + \sqrt{w_S^2 + w_{FR}^2}\right) \tag{1}$$

where

$$w_S = (S - 1.75) \times 0.25 \log(N_5 + 10) \tag{2}$$

for $S > 1.75$, and

$$w_{FR} = 2.18 / N_5^{0.4} (0.4F + 0.6R) \tag{3}$$

where N_5 is the Loudness in sones exceeded for 5 percent of the time, and Sharpness S, Fluctuation Strength F and Roughness R are measured in acum, vacil and asper respectively. For sharpness less than 1.75, the contribution to annoyance from w_S is zero.

OPTIMISATION

A trade-off exists between controlling intelligibility contrast and reducing extraneous noise radiation, which can be perceived as annoying. Two alternative optimisation formulations are available: multi-objective and single-objective optimisation. Both methods require the evaluation of a cost function which in turn involves the synthesis of a masking signal based on a set of parameters, simulation of the sound fields in the bright and dark zones, then evaluation of speech intelligibility in both zones and annoyance in the dark zone.

A common multi-objective optimisation paradigm involves the creation of a *Pareto front* [17], a set of solutions in which the improvement of one parameter,

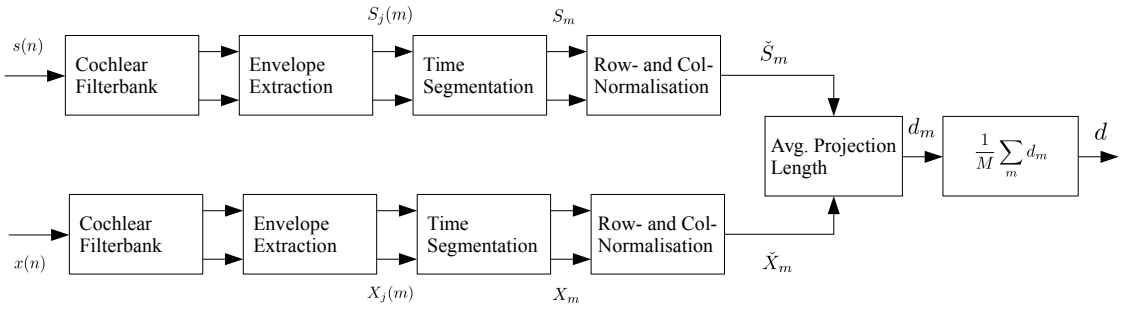

Fig. 6: Block Diagram of ESTOI Algorithm, reproduced from [7]. The clean and noisy signals $s(n)$ and $x(n)$ are passed through a 1/3 octave filterbank, then the temporal envelopes are extracted. The resulting spectrograms $S_j(m)$ and $X_j(m)$ are divided into short-time segments before being normalised in time and frequency. The intermediate indices d_m represent the *distance* between $\check{S}_m$ and $\check{X}_m$. The row- and column-normalisation results in $-1 <= d <= 1$, with a value of $d = 0$ meaning no correlation. In low-noise situations, $x(n) \approx s(n)$, giving values of d close to 1.

say a reduction in annoyance, can only be achieved at the expense of the other parameter, intelligibility contrast. Creating this surface can be computationally demanding as it requires a very large number of cost function evaluations. Optimising for a single objective is conceptually and computationally simpler. Limits of acceptability are placed on one objective, then the parameters are adjusted to minimise the other. Due to the context-dependence of annoyance, it is difficult to place an upper bound on the numerical value of the annoyance metric below which listeners regard signals to be acceptable, despite it being derived from metrics which have well-defined units (sone, acum, vacil, asper). However, intelligibility as reported by the ESTOI algorithm can be mapped to the percentage of words correctly identified in listening tests. This implies that meaningful numerical limits on ESTOI in the bright and dark zones can be chosen, leaving annoyance to be minimised by an optimisation algorithm. Formally, we wish to minimise the Psychoacoustic Annoyance

$$J = PA_{dark} \tag{4}$$

subject to a pair of constraints on the intelligibility in each zone given by

$$\text{ESTOI}_{dark} < \varepsilon_1 \quad \& \quad \text{ESTOI}_{bright} > \varepsilon_2, \tag{5}$$

where ε_1 is the maximum allowable level of intelligibility in the dark zone and ε_2 is the minimum acceptable level of intelligibility in the bright zone. These threshold levels can be set independently for different applications or scenarios, as speech intelligibility is dependent on the familiarity of vocabulary and information content of messages [18]. In this paper, $\varepsilon_1 = 0.3$ and $\varepsilon_2 = 0.6$.

In order to make an informed choice regarding the optimisation strategy and parameter range selection, the cost and constraint functions (Eqs. 4 and 5) can be evaluated with a number of representative test signals.

COST FUNCTION DEPENDENCIES

The following figures show how the output of the ESTOI and Psychoacoustic Annoyance algorithms vary with signal level, using four common masking signals; white noise, pink noise, noise with a power spectrum that matches the speech intended for the bright zone, and multi-talker babble.

Figure 7 shows that almost the full range of objective intelligibility can be achieved with 40 dB of variation in the energy ratio between target sound (speech) and an interferer (TIR). Multi-talker babble and speech-shaped noise are the most effective maskers by this measure as they provide lower levels of intelligibility at the same TIR compared with pink noise and white noise. The trend in this figure can be associated with the negative correlation found between TIR and the

subjective level of distraction that an interfering signal may cause [19]; that is as the intelligibility of a speech signal decreases, the potential for distraction from competing audio increases. In this use case, the distraction of unintended listeners from the material delivered to the bright zone is desirable, whilst the distraction of the intended listener by leakage of the masking signal is undesirable.

Figure 8 shows the relationship between Psychoacoustic Annoyance and masker level for the four base masking signals. Psychoacoustic Annoyance is strongly positively correlated with signal level, which in turn is known to be correlated with loudness. The presence of fifth-percentile loudness in every term of Equation 1 confirms this relationship. The vertical offset between traces can be explained by the different level of sharpness of each signal. Babble is judged to be more annoying than speech-shaped noise due to its fluctuating nature.

Analysis of Figures 7 and 8 together gives information on the likelihood of optimisation constraints being reached. Intelligibility decreases with the target to interferer ratio. Consequently, the intelligibility in the dark zone can be expected to increase as the algorithm minimises annoyance, indicating that in most cases, the active constraint is likely to be the upper limit on dark zone intelligibility. The lower bound on bright zone intelligibility may become significant for arrays with low levels of acoustic contrast between zones.

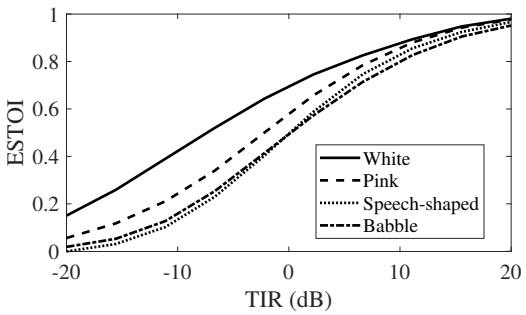

Fig. 7: Variation in the intelligibility (ESTOI) of a recorded sentence corrupted by different masking signals at a range of Target-to-Interferer Ratios (TIR, dB).

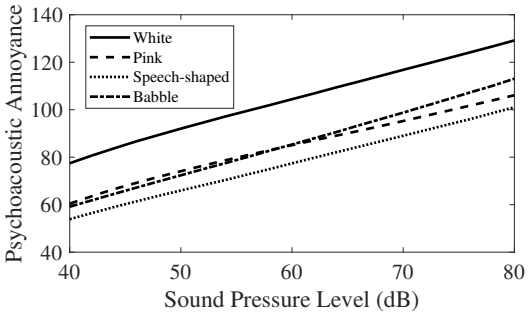

Fig. 8: Variation of Psychoacoustic Annoyance of four common masking signals with SPL. Similarity in gradient is caused by the strong dependence of annoyance on loudness.

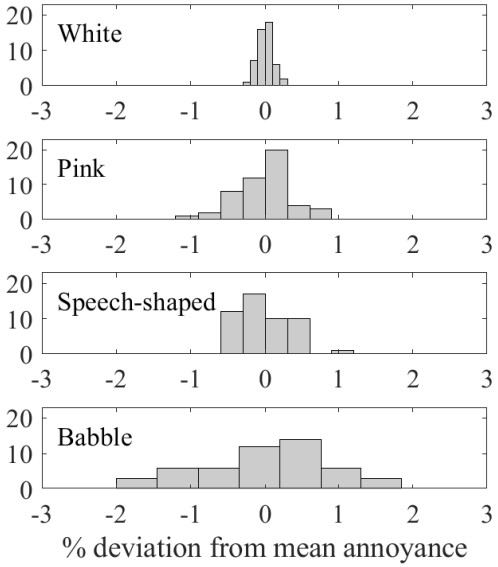

Fig. 9: Percentage deviation from the mean annoyance for 50 ten-second long examples of white noise, pink noise, speech-shaped noise and multi-talker babble, all normalised to 60 dB SPL. Biasing the masking signal spectrum to lower frequencies and increasing roughness and fluctuation strength all increase the uncertainty in the annoyance result.

It is interesting to observe that the roughness and fluctuation strength algorithms are sensitive to the fine structure of their input signals, thus giving a slightly

different output each time the array is simulated, even with statistically similar noises. Histograms of Psychoacoustic Annoyance which show the deviation from the mean for the four input masking signals are presented in Figure 9. This uncertainty means that the cost function J is *stochastic*, rather than deterministic. Standard gradient-based optimisation methods perturb a candidate solution by a small amount in each dimension to determine the next parameter values to test. This method is only guaranteed to find locally optimal points in smooth, convex objective functions. These conditions do not necessarily hold for the simulation represented by Equation 4, and this will be considered in the selection of the optimisation algorithm in the following section.

PATTERN SEARCH

The pattern search algorithm [20] is a gradient-free optimisation method suitable for objective functions $f(x)$ which can be expressed in the form

$$f(x) = E(F(x,\xi)). \tag{6}$$

Where $F(x,\xi)$, corresponds to the results of the array simulation and the expectation $E(\cdot)$ is equal to the underlying objective function $f(x)$. The random variable ξ represents the random variation in the cost function due to the random signals which are generated each time the cost function is called, and the vector of parameter values x is the input to the simulation. The algorithm utilises the assumption that the change in the value of the cost function over a large range of x exceeds any local random variation.

Pattern search can be understood by visualising the parameter space as a multidimensional grid. Each point within the grid has an associated cost, and the extent of the grid, i.e. the maximum and minimum parameter values are set in advance. Starting at an initial point x_0, the pattern search algorithm evaluates the cost function at points in each coordinate direction, spanning a large range of x. The point with the lowest cost then becomes the new starting point, and the pattern search iterates. The size of the pattern is increased after a successful poll, and decreased after an unsuccessful poll, allowing the algorithm to search the function space fully, increasing the likelihood of finding the global minimum. The algorithm halts when the size of the pattern to be searched is smaller than a pre-defined level, here set to correspond with 0.5 dB perturbations in any of the parameters described in Table 1, which is less than a Just Noticeable Difference.

The implementation of the nonlinear constraints on the objective function (Eq. 5) takes advantage of the fact that the pattern search algorithm is a gradient-free method, and is robust to discontinuous cost functions. At a candidate point, if the intelligibility in either zone falls outside of the constraints, the function immediately returns a value of positive infinity, guaranteeing the algorithm will move away from the neighbourhood of these points. A further advantage of this approach is that ESTOI is computationally inexpensive compared to the prediction of annoyance, so no time is wasted computing annoyance at infeasible parametrisations.

OPTIMISATION STRATEGY

It is clear from the results presented in Figures 7 and 8 that signal level is a highly influential parameter on the masking ability and potential annoyance of a signal. Furthermore, the significant difference in ESTOI and Psychoacoustic Annoyance scores between white and pink noise indicate that the spectrum of a potential masking signal also has scope for optimisation. Octave band filters are chosen as the speech frequency range can be covered with six parameters. Critical band filtering would enable more precise spectral control which is better correlated with the response of the ear, but would require 16 parameters to cover the same frequency range, increasing the computational complexity of the optimisation procedure significantly. Table 1 shows the seven parameters that are used to control the masking signal. The initial condition for the optimisation routine sets all parameters at 0 dB, producing a spectrally unmodified masking signal at the same level as the programme.

x_n	Description	Range
x_1	< 125 Hz shelving filter gain	±20 dB
x_2	250 Hz octave band filter gain	±20 dB
x_3	500 Hz octave band filter gain	±20 dB
x_4	1 kHz octave band filter gain	±20 dB
x_5	2 kHz octave band filter gain	±20 dB
x_6	> 4 kHz shelving filter gain	±20 dB
x_7	SPL re. programme level	±20 dB

Table 1: Parameters x_n available for adjustment by the optimisation routine

The filter bandwidths and responses are set to overlap such that if all gain parameters are set at the same value, the signal spectrum remains flat within ± 3 dB. Automatic *make up* gain is applied to the output of the equaliser so that it has the same energy as the input signal. This results in overall sound pressure level control being handled exclusively by the x_7 parameter. x_1 to x_6 can be regarded as controlling the balance of the masking signal's spectrum. This formulation opens the possibility of multiple parametrisations of the same input signal, e.g. two settings where x_1 to x_6 are all shifted by the same value. This could be handled by imposing a constraint on the sum from x_1 to x_6, however the flexibility of allowing the optimisation routine to arbitrarily set parameter values outweighs the potential robustness offered by preventing duplicate parametrisations. In practice, as the pattern search algorithm adjusts each variable in turn, this scenario is rarely encountered.

RESULTS

The results in Tables 2 and 3 show the optimal equalisation and level settings and corresponding values of intelligibility and annoyance respectively. For all four types of base masking signal, the optimisation algorithm produces significant reductions in the predicted annoyance compared to the initial parametrisation. With this array, using unmodified white noise at the same level as the masker would be infeasible as the intelligibility in the dark zone exceeds the limit ε_1. A fortuitous consequence of the optimisation process is the maintenance or slight improvement in bright zone intelligibility observed for all base masking signals except white noise. This exception is not surprising; unmodified white noise does not provide an acceptable level of privacy and thus makes only a small degradation to the bright zone signal.

Figures 10 and 11 show the power spectral density of the masking signals before and after optimisation. Given that the three input random noise samples only differ in their spectra, it is expected that the optimisation routine would adjust the spectra to produce the same optimal signal. The similarity between traces in Figure 11 confirms this expectation. Discrepancies at the lowest and highest octave bands are due to the difference in spectral level of the original signals exceeding the range of adjustment available to the optimisation routine, which is limited to reduce the size of the search space. At speech frequencies, the spectra show good alignment. Further assurance of the convergence of the optimisation routine can be found in Table 3: the annoyance value for optimised white, pink and speech-shaped noise is within the random variation found in Figure 9.

The rightmost column in Table 2 shows the time for the algorithm to converge for each masking signal type. Pink and speech-shaped noise converged in around half the time as white noise, and babble took less than a third of the time, reflecting the size of the change in signal spectrum which the algorithm must effect, particularly at high frequencies, in order to reach the optimum. This shows that pre-shaping of the input noise is advantageous to the algorithm's performance, a characteristic which invites the potential for real-time implementation, as small updates to the masking signal may be rapidly calculated as conditions such as ambient noise or the spectrum of the speech to be masked change over time.

Signal	125Hz	250 Hz	500 Hz	1 kHz	2 kHz	4 kHz	SPL	Time (s)
Initial	0	0	0	0	0	0	0	-
White	20	14	20	16	-4	-20	-4	515
Pink	20	19	20	20	0	-3	-3	222
Shaped	20	1	20	15	9	-1	-1	278
Babble	19	20	20	0	0	0	0	149

Table 2: Equaliser octave band gain settings in dB, Overall sound pressure level of masker relative to programme level, rounded to 1 dB, and algorithm convergence time in seconds. The initial settings represent the starting point given to the pattern search algorithm.

CONCLUSIONS

A method for designing personal audio systems motivated by subjective performance has been developed with a specific focus on improving privacy. The method utilises a synthesised masking signal, which is radiated into the acoustic dark zone to reduce the intelligibility of speech material intended for the bright zone. The masking signal is optimised to reduce the annoyance in the dark zone, a quantity which is estimated using the Psychoacoustic Annoyance metric. For the array

	Signal	ESTOI Dark	ESTOI Bright	PA_{dark}
Initial	White	0.41	0.94	112.8
	Pink	0.20	0.88	96.9
	Shaped	0.18	0.82	79.8
	Babble	0.16	0.84	89.0
Optimised	White	0.30	0.89	74.6
	Pink	0.30	0.88	75.8
	Shaped	0.29	0.87	75.0
	Babble	0.28	0.89	79.8

Table 3: Pre- and post- optimisation cost function results for four masking signals. The pattern search algorithm minimises Psychoacoustic Annoyance in the dark zone, subject to ESTOI less than 0.30 in the dark zone and greater than 0.60 in the dark zone.

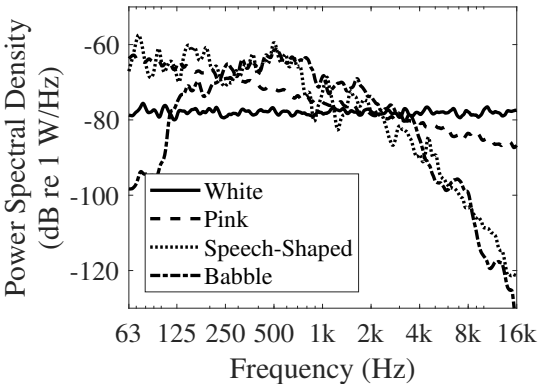

Fig. 10: PSD of base masking signals before optimisation and gain adjustment.

geometry considered in this paper, whose acoustic contrast rises with frequency from around 5 to 15 dB, the most appropriate masker is based on random noise, equalised with a low-pass characteristic rolling off at 1 kHz. Multi-talker babble was predicted to provide similar levels of masking to speech-shaped noise at the same Target to Interferer Ratio, although the inherent fluctuation rendered it subjectively more annoying.

Further work must be carried out into determining adequate values of intelligibility thresholds for different applications. This may require listening tests with a loudspeaker array or a simulated array reproduced over headphones. The former test could also be used to validate the accuracy of the array simulation and to confirm the benefits of the optimisation compared to initial settings.

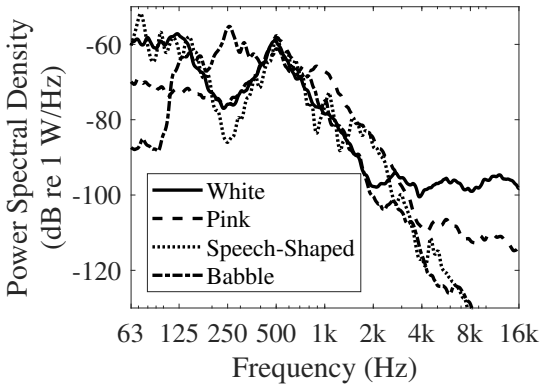

Fig. 11: PSD of optimised masking signals after equalisation and gain adjustment.

ACKNOWLEDGMENTS

This work was supported by an EPSRC Centre for Doctoral Training grant (EP/L015382/1)

References

[1] Druyvesteyn, W. F. and Garas, J., "Personal Sound," *Journal of the Audio Engineering Society*, 45(9), pp. 685–701, 1997.

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

[3] Elliott, S. J. and Jones, M., "An active headrest for personal audio," *The Journal of the Acoustical Society of America*, 119(5), p. 2702, 2006, ISSN 00014966, doi:10.1121/1.2188814.

[4] Cheer, J., Elliott, S. J., and Gálvez, M. F. S., "Design and implementation of a car cabin personal audio system," *AES: Journal of the Audio Engineering Society*, 61(6), pp. 412–424, 2013, ISSN 15494950.

[5] Elliott, S. J., Cheer, J., Murfet, H., and Holland, K. R., "Minimally radiating sources for personal audio," *The Journal of the Acoustical Society of*

America, 128(4), pp. 1721–1728, 2010, ISSN 0001-4966, doi:10.1121/1.3479758.

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

[7] Jensen, J. and Taal, C. H., "An Algorithm for Predicting the Intelligibility of Speech Masked by Modulated Noise Maskers," *IEEE/ACM Transactions on Audio Speech and Language Processing*, 24(11), pp. 2009–2022, 2016, ISSN 23299290, doi:10.1109/TASLP.2016.2585878.

[8] Fastl, H. and Zwicker, E., *Psychoacoustics: Facts and models*, Springer, 2007, ISBN 3540231595, doi:10.1007/978-3-540-68888-4.

[9] Elliott, S. J., Cheer, J., Choi, J.-W., and Kim, Y., "Robustness and Regularization of Personal Audio Systems," *IEEE Transactions on Audio, Speech and Language Processing*, 20(7), pp. 2123–2133, 2012.

[10] P.862, "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," Standard, International Telecommunications Union, 2001.

[11] ANSI/ASA S3.5-1997, "ANSI/ASA S3.5-1997 - Methods for Calculation of the Speech Intelligibility Index," Standard, ANSI, 1997.

[12] Taal, C. H., Hendriks, R. C., Heusdens, R., and Jensen, J., "An Algorithm for Intelligibility Prediction of Time-Frequency Weighted Noisy Speech," 19(7), pp. 2125–2136, 2011, ISSN 1558-7916, doi:10.1109/TASL.2011.2114881.

[13] P.563, "Single-ended method for objective speech quality assessment in narrow-band telephony applications," Standard, International Telecommunications Union, 2004.

[14] Steeneken, H. J. M. and Houtgast, T., "A physical method for measuring speech transmission quality," *The Journal of the Acoustical Society of America*, 67(1), pp. 318–326, 1980, ISSN 0001-4966, doi:10.1121/1.384464.

[15] European Parliament and Council of the European Union, "Assessment and management of environmental noise (EU Directive)," *Official Journal of the European Communities*, (L189), pp. 12–25, 2002, ISSN 0959-6526, doi:10.1016/j.jclepro.2010.02.014.

[16] McGuire, S. and Davies, P., "An Overview of Methods To Quantify Annoyance Due To Noise With Application To Tire-Road Noise," (February), 2008.

[17] Miettinen, K., *Nonlinear Multiobjective Optimization*, Springer Science and Business Media, 1999.

[18] IEC 60268-16:2011, "IEC 60268-16:2011 Sound system equipment - Part 16: Objective rating of speech intelligibility by speech transmission index," Standard, International Electrotechnical Commission, 2011.

[19] Rämö, J., Bech, S., and Jensen, S. H., "Real-Time Perceptual Model for Distraction in Interfering Audio-on-Audio Scenarios," *IEEE Signal Processing Letters*, 24(10), pp. 1448–1452, 2017, ISSN 1070-9908, doi:10.1109/LSP.2017.2733084.

[20] Abramson, M. A., *Pattern Search Filter Algorithms for Mixed Variable General Constrained Optimization Problems*, Ph.D. Thesis, Rice University, 2002.