Combining Artificial and Natural Background Noise in Personal Audio Systems

Daniel Wallace
Institute of Sound and Vibration Research
University of Southampton
Southampton, UK
Email D.Wallace@soton.ac.uk

Jordan Cheer
Institute of Sound and Vibration Research
University of Southampton
Southampton, UK
Email J.Cheer@soton.ac.uk

Abstract—**Personal audio systems are designed to deliver spatially separated regions of audio to individual listeners. This paper presents a method for improving the privacy of such systems. The level of a synthetic masking signal is optimised to provide specified levels of intelligibility in the bright and dark sound zones and reduce the potential for annoyance of listeners in the dark zone by responding to changes in ambient noise. Results from a simulated personal audio system indicate that less acoustic contrast is required to produce the same level of privacy when artificial masking is included in the system design, compared with relying on the masking effect of background noise alone. As privacy requirements become more challenging, the advantage gained by incorporating artificial masking increases.**

I. 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 contrasting zones of sound, conventionally designated as acoustically *bright* and *dark*. However, in this work, the contrast is defined as the inter-zone difference in speech intelligibility, rather than as a measure of the difference in energy between the two zones.

Personal audio systems are restricted in directivity due to practical limits on the number of loudspeakers in the array. This can lead to distraction or annoyance of nearby people, or a lack of privacy as programme material may remain intelligible outside of the bright zone. The proposed personal audio system undertakes to restore privacy by using the array to radiate 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 limiting its intrusive effect on the programme in the bright zone.

For such a system to be successful, it is not sufficient to solely optimise the masking signal to provide privacy for the listener in the bright zone, as high masker levels in the dark zone demanded by this process may be regarded as noise pollution. This line of enquiry, i.e. consideration of the perceptual relevance of sound leaking from one zone into another, has been taken by contributors to the POSZ Project [7] to inform loudspeaker positioning, [8], [9] and the choice of sound zoning methods [10]–[12]. A recent article by Donley et. al. [13] has demonstrated with experimentally validated simulations that the addition of secondary masking can improve speech intelligibility contrast between zones, and that optimisation techniques can be used to further hone this performance by adjusting the masker's spectrum. The optimisation procedure used in the present work differs from the formulation in [13] by placing constraints on intelligibility in the bright and dark zones, and minimising a metric correlated with the sensation of annoyance [14] in the dark zone, all whilst explicitly including the effect of background noise in simulations. The present paper shows an extension to previous work by the authors [15] in which the level and spectrum of the masker are optimised to minimise dark zone annoyance subject to constraints on intelligibility. Here, we take into account the impacts, both positive and negative, that background noise may have on systems which employ secondary masking signals.

In a realised personal audio system, the number and positions of the loudspeakers and microphones used in the array and audio zones have a pronounced effect on the frequency dependent level of acoustic contrast that may be achieved. Likewise, the choice of zoning method may affect acoustic contrast levels, alongside subjective factors such as target quality [16]. However, in order to facilitate general conclusions on the methods discussed in this paper, the sound fields in the bright and dark zones are simulated by specifying a frequency independent level of acoustic contrast, and summing the programme, masker and background noise after setting their respective sound pressure levels. This is a significant and necessary simplification which first investigates the principle difference between sound zones of this type, that is, the relative levels of programme, masker and natural background noise. Future developments to this approach may encapsulate further subtleties inherent to the production of sound zones, such as the effects of frequency dependent acoustic contrast, alternative zoning methods, or spatial aliasing.

Information about the performance of the system is ascertained through the use of metrics which provide a mapping

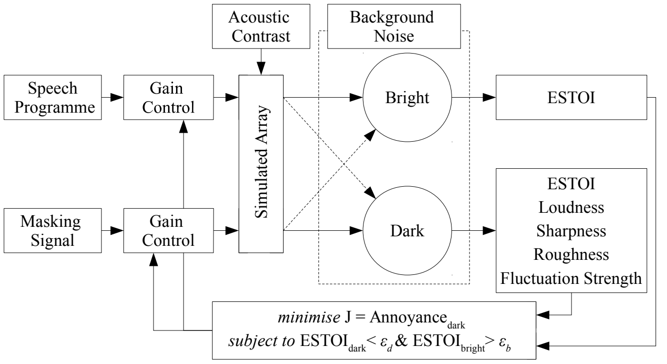

Fig. 1. Block Diagram of the proposed personal audio system simulation. Three signals are input into the model: speech programme intended for the bright zone, a masking signal for the dark zone, and a representative background noise signal. This signal is added to both zones, which are separated by a fixed level of Acoustic Contrast. The level of the programme and masker are updated by a constrained optimisation loop that minimises the estimated annoyance in the dark zone whilst maintaining a minimum level of intelligibility (ESTOI) contrast set by ε_d and ε_b.

between objective features of a signal and expected responses from a population. The optimisation relies on the Extended Short-Time Objective Intelligibility (ESTOI) [17] metric due to its wide applicability to both steady state and time-varying maskers and Psychoacoustic Annoyance [14] which 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. Hereafter, the *performance* of a personal audio system refers to the ability to provide a bright zone where a speech signal is adequately intelligible, and a dark zone in which the speech signal is sufficiently unintelligible, and the potential for annoyance by the masker is minimised.

The potential influence that background noise might have on the performance of a personal audio system is discussed in Section II. This is followed by the results of two sets of simulations: Section III shows the process of determining the combination of programme and masker signal levels that will simultaneously minimise Psychoacoustic Annoyance and satisfy intelligibility constraints, and Section IV shows how these optimal signal levels vary with different levels of acoustic contrast and specified intelligibility limits. The paper concludes with some practical considerations that must be taken into account when implementing such a system.

II. The Effects of Background Noise

The well known strong positive correlation between signal to noise ratio and speech intelligibility is one of the first considerations when designing conventional sound reinforcement systems [18]. A public address or voice alarm system must overcome ambient noise so that messages can be clearly understood. The proposed personal audio system must provide this functionality in a confined spatial region, with the opposite goal applicable elsewhere; here, background noise may be used advantageously to mask unwanted speech.

Research into the physiological process of understanding speech in noise has kept pace with the technical development of audio products. Technology has been used to both improve and selectively reduce the intelligibility of speech under different background noise conditions. To name two common examples, sound masking in open-plan offices aims to reduce distraction from neighbouring colleagues by increasing ambient noise [19], whilst directional microphone arrays have been designed to improve the intelligibility of speech for users of hearing aids [20]. Currently, technology used to produce localised regions of sound has seen commercial success in museum and trade show exhibits, or for targeted advertising [21], [22]. The public nature of these spaces highlight the susceptibility of personal audio systems to background noise interference. Systems could be used to relay spoken communication through a security screen, such as between a bank teller or pharmacy clerk and customers, necessitating a focus on privacy.

The primary motivation for considering the inclusion of a masking signal is to maximise the performance of a given array configuration. Without a dedicated masking signal, the system must rely upon the acoustic contrast control offered by the array and the masking effect of any ambient background noise to reduce intelligibility in the dark zone.

A masking signal may be added at no additional hardware cost to deliver potentially significant performance improvements in terms of intelligibility contrast. Equivalently, the performance in this respect of a high power array with many loudspeaker elements may be achievable by a smaller array which uses the proposed method, reducing cost.

In order to test the feasibility of using a masking signal to improve the performance of a personal audio system, a number of configurations are simulated. These investigations are detailed in the next section.

III. Intelligibility and Annoyance Surfaces

The output of the ESTOI algorithm [17] can be mapped to the results of intelligibility scores obtained in listening tests, often quantified in terms of the percentage of words recognised. This implies that meaningful numerical limits on ESTOI in the bright and dark zones can be chosen, given information about the context and familiarity of words in messages [23], which strongly affects the ability to extract meaning from a sentence [24]. For the purposes of the tests in this section, the intelligibility of programme material in the dark zone may not exceed $\varepsilon_d = 0.2$, representing a degree of privacy where speech remains audible, but would require considerable effort to understand, and the intelligibility in the bright zone may not be less than $\varepsilon_b = 0.6$, where speech may be clearly understood.

In the following simulation, the array is set to provide an acoustic contrast level of 10 dB between zones. The programme material is a recorded sentence from the Harvard sentence corpus spoken by a male speaker [25]; the masking signal is random noise equalised to match the power spectrum of the programme; and background noise is taken from a recording of a supermarket checkout area [26], reproduced at a level of 60 dB SPL. Single representative samples of speech programme material and artificial masking are varied in level relative to background noise, as averaging over a range of samples to overcome the variability in intelligibility prediction with different spoken sentences would carry a significant computational cost.

Simulation results are presented in three coloured contour plots. Figures 2 and 3 show contours of intelligibility in the bright and dark zones respectively, with various programme and masking signal levels measured relative to the background noise. Intuitively, increasing the programme level and reducing the masking level, that is moving towards the upper left corner of each plot, results in increased predicted intelligibility in both zones. The two plots may be used in conjunction to find a feasible region where both intelligibility constraints, $\mathrm{ESTOI}_{\mathrm{bright}} > 0.6$ and $\mathrm{ESTOI}_{\mathrm{dark}} < 0.2$ are met, a region bounded by the heavy dashed and dotted contours from Figures 2 and 3. This region is replicated in Figure 4, where the colour scale here indicates annoyance as predicted by the Psychoacoustic Annoyance metric [14]. Despite the metric having well-defined numerical units, inherited from its derivation from Loudness, Sharpness, Roughness and Fluctuation Strength, the actual experience of noise annoyance is highly dependent on context [27], restricting the comparisons between annoyance ratings that can be justified. However, as the metric is designed to be monotonically related to the experience of annoyance, and the dark zone sound fields represented by points in the contour plot are not contextually dissimilar, it is reasonable to use Psychoacoustic Annoyance in this case as a target for minimisation.

The minimum point within the feasible region is found at the junction between the boundaries where $\mathrm{ESTOI}_{\mathrm{bright}} > 0.6$ and $\mathrm{ESTOI}_{\mathrm{dark}} < 0.2$, with the programme 3 dB above the background level and the masker 4 dB below background. Interestingly, the region where the highest value of the annoyance metric is predicted is at the upper-left of Figure 4, where the dominating sound source is the programme material which has bled from the bright zone into the dark zone. This high value of annoyance is predicted due to greater fluctuation strength of speech compared to the masking signal and background noise.

The ESTOI metric processes signals linearly, and so is only sensitive to changes in the ratio between programme, masker and background levels, rather than absolute levels. The dependence of Psychoacoustic Annoyance on a loudness model means that the nonlinearity of the human auditory system is captured. Therefore, a small difference in the shape of the annoyance contour plot can be expected at different background levels, though the general trend is unaffected. This trend yields the conclusion that for a given level of acoustic contrast, predicted annoyance is minimised when signals can be chosen to exactly meet the constraints on intelligibility. The next section describes investigations into the minimum level of acoustic contrast for which there exists such a feasible region, and the maximum level of acoustic contrast for which the reproduction of a masking signal becomes unnecessary.

IV. Acoustic Contrast Selection

In order to satisfy a particular performance requirement, a personal audio system must produce sufficient acoustic contrast to constrain the programme and masking signals to their respective zones. The previous section showed that for a system providing 10 dB of acoustic contrast, the combination of programme and masking signal levels which satisfies both of these constraints provides minimal Psychoacoustic Annoyance in the dark zone. This section shows how these optimal signal levels depend on the level of acoustic contrast provided by the array.

As previously, the bright and dark sound fields are produced by simulating frequency independent levels of acoustic contrast from 5 to 14 dB, with background noise fixed at 60 dB SPL. The intelligibility score in each zone is calculated by the ESTOI algorithm, the results from which are used by an optimisation algorithm to independently adjust the programme and masker levels. The cost function of this optimisation process reaches a global minimum when the intelligibility

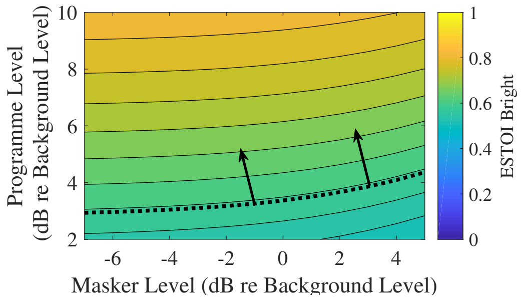

Fig. 2. Contour plot of bright zone intelligibility (ESTOI) with variation in masker and programme levels, measured relative to the background level of 60 dB SPL. The dotted contour indicates an ESTOI of 0.6. The region above this line, indicated with arrows, represents programme and masker combinations that provide sufficient intelligibility in the bright zone.

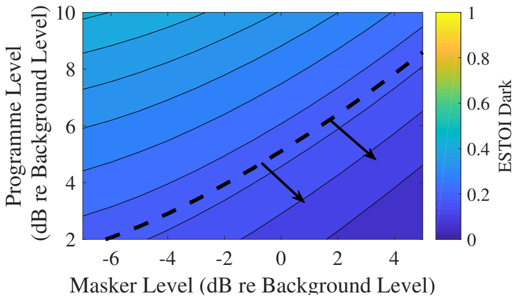

Fig. 3. Contour plot of dark zone intelligibility (ESTOI) with variation in masker and programme levels, measured relative to the background level of 60 dB SPL. The dashed contour indicates an ESTOI of 0.2. The region below this line, indicated with arrows, represents programme and masker combinations for which privacy is achieved.

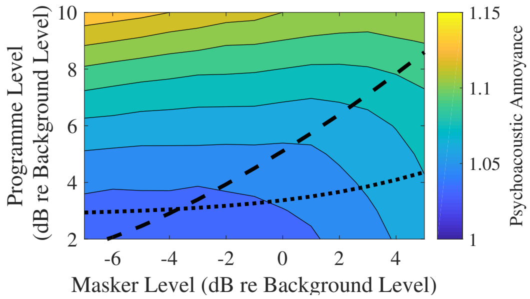

Fig. 4. Contour plot of estimated Psychoacoustic Annoyance in the dark zone, with variation of programme and masker levels. The dotted and dashed lines represent bright zone and dark zone intelligibility limits from Figures 2 and 3. The region between the lines satisfies both limits.

constraints in both zones are just satisfied, as dark zone annoyance is minimised.

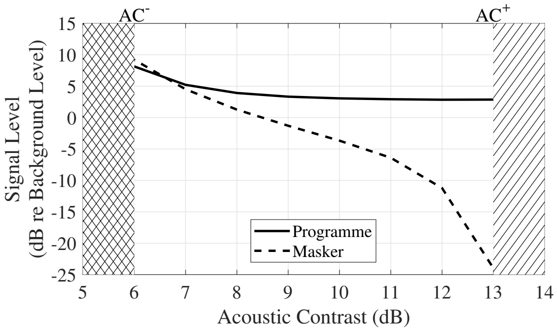

Fig. 5. Programme and masker levels relative to the background noise level to achieve ESTOI$_{bright} = 0.6$ and ESTOI$_{dark} = 0.2$. The acoustic contrast in the cross-hatched region below AC^- is too low to achieve the performance standard with any programme and masker level combination. In the diagonally hatched region above AC^+ an artificial masking signal is not necessary as the background noise produces sufficient masking.

TABLE I
Acoustic Contrast levels required to achieve a level of performance given by ε_d and ε_b with additional masking (AC^-) or without additional masking (AC^+).

AC^- (dB)			ε_d		AC^+ (dB)			ε_d	
		0.1	0.2	0.3			0.1	0.2	0.3
	0.6	8	6	4		0.6	18	13	9
ε_b	0.7	9	7	6	ε_b	0.7	21	16	12
	0.8	11	9	7		0.8	24	19	15

Figure 5 shows the optimal programme and masker signal levels for each value of acoustic contrast tested. In the cross-hatched region, no valid solution for the optimisation is found, as the acoustic contrast is too low to provide the required intelligibility contrast; any increase in the programme level would unacceptably raise dark zone intelligibility, and any increase in the masking signal level would result in excessive degradation of the programme signal in the bright zone. At $AC^- = 6$ dB, a feasible pair of signals is found; the energy of both signals is 7-9 dB greater than the background level, potentially raising dark zone annoyance compared to designs with more acoustic contrast and lower signal levels. At higher levels of acoustic contrast, the programme level plateaus at 3 dB above background and the optimal masking signal level decreases as the contribution of the background noise to obscuring the programme in the dark zone becomes more significant. With the tested combination of signals, acoustic contrast levels in excess of $AC^+ = 13$ dB, marked with diagonal hatching, provide sufficient separation between zones for the masking signal to be omitted entirely. Increasing acoustic contrast further and maintaining programme level will result in improved privacy in the dark zone; alternatively, the programme signal may be increased to improve bright zone intelligibility while maintaining the previously set intelligibility limit in the dark zone.

Table I shows the variation of AC^- and AC^+ with different intelligibility limits in each zone. These values show the level of frequency independent acoustic contrast required to satisfy intelligibility constraints with minimal dark zone annoyance, with (AC^-) or without (AC^+) additional masking. As higher performance is demanded by increasing ε_b and decreasing ε_d, more acoustic contrast is required, but the advantage gained by introducing additional masking, indicated by the difference between AC^- and AC^+ also increases.

V. Conclusions

A series of simulations have been presented to show that by combining an artificial masking signal with natural background noise, a personal audio system can be designed to provide bright and dark audio zones with contrasting levels of speech intelligibility. This is quantified by stipulating a maximum level of intelligibility in the dark zone, preserving privacy, and a minimum level of intelligibility in the bright zone, protecting the programme material from degradation.

The acoustic contrast offered by a personal audio system may be limited in order to improve robustness, increase dynamic range, or save cost by reducing the number of array elements. This in turn would reduce the maximum level of intelligibility contrast achievable by the system. Reproducing a variable level masking signal in the dark zone of the array may improve intelligibility contrast beyond that possible by traditional designs, or may allow for a more robust, cost-effective array design to be chosen, if a particular performance target is set.

In order to implement a system of this type, care must be taken over the choice of ε_b and ε_d. These limits are specific to the particular application the system is intended for, so may be determined by conducting listening tests with potential users of the system. Further practical considerations must be made for systems situated in public places, as microphones may have to be incorporated to compensate for a wide range of background noise levels.

Acknowledgments

We are grateful for the referees for their constructive input. This work was supported by an EPSRC Doctoral Training Centre grant (EP/L015382/1).

References

[1] W. F. Druyvesteyn and J. Garas, “Personal Sound,” *Journal of the Audio Engineering Society*, vol. 45, no. 9, pp. 685–701, 1997.

[2] 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,” *Journal of the Acoustical Society of America*, vol. 125, no. 4, pp. 2091–2097, 2009.

[3] S. J. Elliott and M. Jones, “An active headrest for personal audio,” *The Journal of the Acoustical Society of America*, vol. 119, no. 5, p. 2702, 2006.

[4] J. Cheer, S. J. Elliott, and M. F. S. Gálvez, “Design and implementation of a car cabin personal audio system,” *AES: Journal of the Audio Engineering Society*, vol. 61, no. 6, pp. 412–424, 2013.

[5] 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*, vol. 128, no. 4, pp. 1721–1728, 2010.

[6] J.-W. Choi and Y.-H. Kim, “Generation of an acoustically bright zone with an illuminated region using multiple sources,” *Journal of the Acoustical Society of America*, vol. 111, no. 4, pp. 1695–1700, 2002.

[7] University of Surrey, “Perceptually Optimised Sound Zones,” 2018. [Online]. Available: http://iosr.surrey.ac.uk/projects/POSZ/index.php

[8] 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*, vol. 19, 2013. [Online]. Available: http://asa.scitation.org/doi/abs/10.1121/1.4800873

[9] M. Olik, P. J. B. Jackson, P. Coleman, and J. A. Pedersen, “Optimal source placement for sound zone reproduction with first order reflections,” *The Journal of the Acoustical Society of America*, vol. 136, no. 6, pp. 3085–3096, 2014. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.4898423

[10] P. Coleman, P. J. Jackson, M. Olik, M. Olsen, M. Moller, and J. A. Pedersen, “The influence of regularization on anechoic performance and robustness of sound zone methods,” *Proceedings of Meetings on Acoustics*, vol. 19, no. 1, pp. 055 055–055 055, 2013. [Online]. Available: http://scitation.aip.org/content/asa/journal/poma/19/1/10.1121/1.4799031

[11] P. Coleman, P. J. B. Jackson, M. Olik, M. Møller, M. Olsen, and J. Abildgaard Pedersen, “Acoustic contrast, planarity and robustness of sound zone methods using a circular loudspeaker array,” *The Journal of the Acoustical Society of America*, vol. 135, no. 4, pp. 1929–1940, 2014. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.4866442

[12] P. J. Jackson, F. Jacobsen, P. Coleman, and J. Abildgaard Pedersen, “Sound field planarity characterized by superdirective beamforming,” *Proceedings of Meetings on Acoustics*, vol. 19, no. 1, p. 055056, 2013.

[13] 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*, vol. 26, no. 4, pp. 1–15, 2018.

[14] H. Fastl and E. Zwicker, *Psychoacoustics: Facts and models*. Springer, 2007.

[15] D. Wallace and J. Cheer, “Optimisation of Personal Audio Systems for Intelligibility Contrast,” in *Proc. 144th Audio Engineering Society Convention*. Audio Engineering Society, 2018. [Online]. Available: https://eprints.soton.ac.uk/420254/

[16] T. Betlehem, W. Zhang, M. A. Poletti, and T. D. Abhayapala, “Personal Sound Zones,” *IEEE Signal Processing Magazine*, vol. March, pp. 81–91, 2015.

[17] J. Jensen and C. H. Taal, “An Algorithm for Predicting the Intelligibility of Speech Masked by Modulated Noise Maskers,” *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 24, no. 11, pp. 2009–2022, 2016.

[18] J. Lochner and J. Burger, “The Intelligibility of Re-inforced Speech,” *Acustica*, vol. 9, pp. 31–38, 1959. [Online]. Available: http://ci.nii.ac.jp/naid/110008109706/

[19] A. Haapakangas, E. Kankkunen, V. Hongisto, P. Virjonen, D. Oliva, and E. Keskinen, “Effects of five speech masking sounds on performance and acoustic satisfaction. implications for open-plan offices,” *Acta Acustica united with Acustica*, vol. 97, no. 4, pp. 641–655, 2011.

[20] G. H. Saunders and J. M. Kates, “Speech intelligibility enhancement using hearing-aid array processing.” *The Journal of the Acoustical Society of America*, vol. 102, no. 3, pp. 1827–37, 1997. [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/9301060

[21] Directional Audio, “Case Studies,” 2018. [Online]. Available: http://www.directionalaudio.co.uk/case-studies

[22] OgilvyNewZealand, “All Good Bananas - Listen to your conscience,” 2011. [Online]. Available: https://www.youtube.com/watch?v=cef4DDQ-CEc

[23] British Standards Institution, “Sound system equipment - Part 16: Objective rating of speech intelligibility by speech transmission index, Annex G,” British Standards Institution, Standard, sept 2011.

[24] A. Boothroyd and S. Nittrouer, “Mathematical treatment of context effects in phoneme and word recognition,” *The Journal of the Acoustical Society of America*, vol. 84, no. 1, pp. 101–114, 1988. [Online]. Available: http://asa.scitation.org/doi/10.1121/1.396976

[25] IEEE, “Standards Downloads,” 2017. [Online]. Available: standards.ieee.org/downloads/269

[26] SOUND and IMAGE FX, “Supermarket(grocery store) cashier area ambient sound effect,” 2015. [Online]. Available: https://www.youtube.com/watch?v=sawtjWDCm7I

[27] R. Lyon, *Designing for Product Sound Quality*, 1st ed. New York: Marcel Dekker Inc., 2000.