Effect of adding artificial reverberation to speech-like masking sound

Yusuke Hioka*, Jen W. Tang*, Jacky Wan*

Department of Mechanical Engineering, University of Auckland, Auckland 1142 New Zealand

Abstract

Time-reversed speech has been known to effectively mask information for speech privacy applications. However, the annoyance and distraction caused by the time-reversed speech-like masking sound is higher than other masking sound. This study investigates the effects of adding artificial reverberation to the time-reversed speech. Subjective listening tests have been conducted to measure the intelligibility of target speech, annoyance and distraction caused by the masking sound. The experimental results suggest that adding artificial reverberation to a speech-like masking sound has a significant effect to reduce the annoyance level while maintaining the masking effectiveness of the original masking sound. A trend was also observed that the addition of artificial reverberation could reduce the level of distraction caused by the masking sound.

Keywords: speech privacy, masking, room reverberation, intelligibility, distraction, annoyance

1. Introduction

Problems arising from the acoustical privacy point of view [1] in public spaces have been known to be an issue, especially in highly populated cities where people are inevitably sharing limited spaces with one another. Due to the lack of acoustical privacy has been known to affect the human's health both physically and psychologically [2], keeping the acoustical privacy in public spaces will significantly reduce social loss. Installing physical structures that reduce the energy of sound reaching the unintended listeners, e.g. installing partition boards or walls that acoustically separate the space of the unintended listeners, may solve the problem. However, installing such structures is often practically infeasible due to space constraint and is also detrimental in spaces where their *openness* is sought such as open plan offices.

Masking is the most commonly used technique to make a target speech unintelligible to the unintended listeners without needing to install any physical structures [3, 4, 5, 6, 7, 8, 9, 10]. This is achieved by projecting a jammer sound (the masking sound) into the area where the unintended listeners are located. Since the early days of sound masking systems, an extensive range of masking sound have been used and studied for their effectiveness in reducing the intelligibility of the target speech. The commonly used masking sounds today are stationary noise (e.g. white noise, pink noise, HVAC (Heating, Ventilating, and Air Conditioning) system's noise [6]) and natural sound (e.g. rain noise, river noise, babble noise). Although these masking sounds, especially with natural sounds, have been said to help boost human emotions and improving cognitive abilities [11], these sounds are only effective enough to render speech unintelligible when the volume of the target speech is below a certain threshold (i.e. very low target-to-masker ratio (TMR)). Research has therefore been ongoing into finding a more efficient masking sound such as speech-like signals, which is also known as *informational masking* [12].

One of the known effective speech-like masking sound is the processed-target speech [3, 4, 13, 14]. Due to the similar spectral envelope between the masking sound and the target speech, the processed-target speech use as a masking sound will degrade the intelligibility of the target speech more efficiently. A mixture of this signal and a stationary noise has also been studied [6]. Some studies have reported [4, 13, 14] that using time-reversed signal of the target speech is more efficient in reducing speech intelligibility. However, the study [14] also concluded that the time-reversed speech causes annoyance and distraction to listeners in return for its ef-

*Corresponding author

Email address: yusuke.hioka@ieee.org (Yusuke Hioka)

ficiency. Hence the design of another masking sound which maintains its masking efficiency while minimising the annoyance and distraction to listeners has been still an open problem.

This study explores a solution to compromise the suggested problem by adding a reverberant effect to a speech-like masking sound. According to the discussion in [14] the distraction and annoyance may be caused by two facts; one is the intelligibility and another is the variability of intensity of the masking sound. It can be hypothesised that the distraction and annoyance may be mitigated by reducing these two aspects in the masking sound by applying signal processing. Generally, reverberation is known to be detrimental to speech intelligibility [15, 16]. Although the time-reversed speech itself has already lost its original context of the speech, it still sounds like a speech and draws an attention of listeners. The proposed approach aims to make the masking sound less attractive by reducing the intelligibility of the sound by adding reverberation. Meanwhile, from signal processing point of view, adding reverberation is equivalent to convolving an impulse response of a reverberant room to the original masking sound. Since such room impulse response often plays a role as a low-pass or band-pass filters, the pulsive part (i.e. signal components in high frequency) of masking sound will be removed which would also contribute to mitigate the negative effects of the speech-like masking sound. This study investigates the effect of adding artificial reverberation to the speech-like masking sound by measuring the intelligibility, distraction and annoyance through subjective hearing tests.

The rest of this paper is organised as follows. Section 2 discusses the signal processing method to develop the masking sound to which the artificial reverberation is added. Methodologies for the subjective listening tests to measure the key three aspects of the proposed masking sound are introduced in Section 3, which is followed by their results and discussion in Section 4. Finally the paper is concluded with some remarks in Section 5.

2. Design of reverberant masking sound

Figure 1 shows the process to generate the masking sound with reverberation, the details of which will be discussed in this section.

2.1. Time-reversed speech as masking sound

Over the years, many research focusing on the effect of the speech-like signals have been conducted to find an effective masking sound. From these research

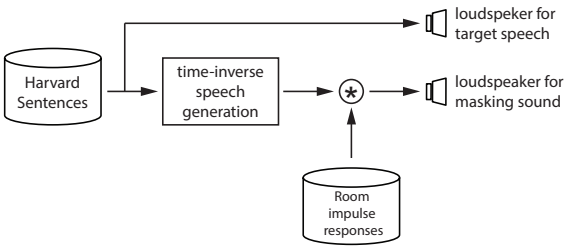


Figure 1: Masking sound generation process.

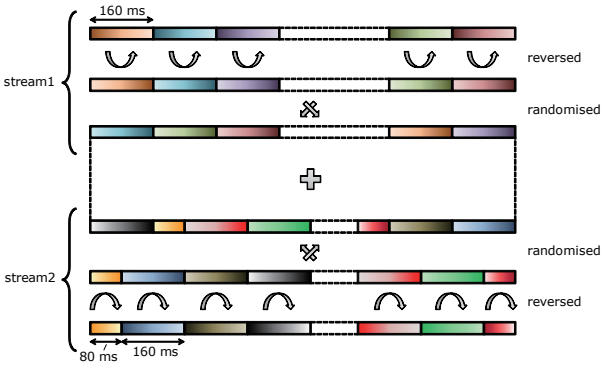


Figure 2: Generation of time-reversed speech [4].

findings thus far, the masking sound using the time-reversed speech has been concluded to be one of the most effective speech-like masking sound in terms of reducing speech intelligibility level but is also deemed to be distracting and annoying to the listeners [4]. This study also employs a time-reversed speech as the *seed* of masking sound and investigates the effect of adding reverberation to the masking sound to overcome the distraction and annoyance problems while maintaining its core speech masking effectiveness.

To generate a time-reversed speech the procedure presented in [4] is followed. An original speech file is first replicated into two identical streams; in which the first stream is split into frames of 160 ms long, while the second stream is split into frames of 160 ms after the first 80 ms of the speech signal. Once completed, these 160 ms frames of both sound streams are reversed and are then randomly swapped against one another in each of the streams. Finally, both streams are added together to form a complete time-reversed speech signal. The procedure is illustrated in Fig. 2.

2.2. Implementation of artificial reverberation

In order to add reverberation effect to the masking sound, an artificially generated room impulse response is convolved with the time-reversed speech. The

methodology employed to produce such a room impulse response (RIR) is the image source method (ISM) [17], which has been employed in various researches in acoustical signal processing to simulate the RIR of a shoebox room. The improved algorithm of the ISM by Lehmann *et al.* [18], which is available via an open-source Matlab code, is deployed in this study. A speech-like masking sound with a room reverberation effect embedded is then generated by

$$y(t) = h(t) * s(t), \tag{1}$$

where $*$ denotes convolution and $h(t)$ and $s(t)$ are the signals of the generated RIR and an arbitrary time-reversed speech, respectively.

2.3. Reverberation intensity

To change the intensity of the reverberation effect added onto the masking sound, different sets of RIRs have to be implemented to the same signal. A key scope of this study is to observe how much reverberation added to speech-like maskers can affect the overall speech intelligibility of the target speech caused by the masking sound while maintaining a low distraction and annoyance level. The RIRs are generated according to the amount of reverberation to be added to a masking sound measured by the direct-to-reverberation ratio (DRR) [19]. In this study, the DRR is defined by

$$\text{DRR [dB]} = 10\log_{10}\left(\frac{\sum_{t=t_d-t_0}^{t_d+t_0}|h(t)|^2}{\sum_{t=0}^{t_d-t_0}|h(t)|^2 + \sum_{t=t_d+t_0}^{\infty}|h(t)|^2}\right) \tag{2}$$

where t_d is the time instance when the direct signal arrives. t_0 is set to 8 ms according to [20].

In the ISM, a RIR is specified by the following parameters: i) dimension of the room, ii) source position, iii) receiver position, and iv) the reflection coefficient of walls, ceiling and floor, all of which affect the DRR of the generated RIR. Out of these parameters in this study, the reflection coefficient is varied while all the other parameters are set to fixed values in order to generate a RIR with a specified DRR. For simplicity the same reflection coefficient is assumed for every wall, ceiling and floor.

3. Methodology for listening tests

3.1. Stimuli

A speech in the database of Harvard sentences [21] was randomly selected and utilised for the target speech to be masked by the masking sound as well as for the seed speech to generate the reverberant masking sound. The Harvard sentences consist of phonetically balanced sentences that use specific phonemes in the same frequency seen in the English language.

In total, seven different masking sounds were tested in the experiment, which included four time-reversed masking sounds with reverberation of various DRRs and three existing masking sound for comparison. The selected DRRs for generating the reverberant time reversed masking sound were 6, 0, −6, and −8 dB whose corresponding reverberation times [22] are summarised in Table 1. Initially, both DRRs of −12 dB and −10 dB were also considered as part of the DRR subsets. However, both these DRRs were omitted due to their high reflection coefficients produced an unnatural standing wave-like effect, which was found to affect the annoyance in a preliminary testing. The other parameters for the ISM to simulate the RIRs are summarised in Table 2.

For the existing masking sounds, the original (i.e. non-reverberant) time reversed speech (`t-rev`), pink noise (`pink`) and their mixture (`mix`) were employed. The energy ratio of the pink noise to the time reversed speech to generate the mixture (`mix`) was ranging from 3.2 to 3.7 (i.e. 5.1 – 5.7 dB). All masking sounds except the pink noise were generated from exactly the same sentence selected as the target speech. All of these masking sounds were normalised by their power in order to keep the target-masker ratio (TMR) at the listener's position consistent. This was made possible by having an audio file in stereo, where one track contains the masking sound and another track contains the target speech, which allows both tracks to play at the same time but at different loudspeakers. Due to the time constraint that each listener can spend on the listening test, only 0 dB was chosen for the TMR as it has shown reasonable speech masking effects in a previous study [14]. The sampling rate of the audio files was 16 kHz.

3.2. Testing environment

The testing was conducted in the Listening Room at the University of Auckland Acoustics Laboratory [23], the general layout of which is shown in Figure 3. This section summarises the acoustics and device setup in the listening room.

3.2.1. Acoustics

The original reverberation time T_{60} [22] of this listening room was initially measured to range between 0.5 to 0.7 s [23] at the frequency between 100 and 1 kHz where ordinary speech energy is dominant. Because this

Table 1: List of masking sound used in the experiment

ID	Masking sound	DRR	T_{60} (ms)
`Reverb_6`	time reversed speech with reverberation	6 dB	80 ms
`Reverb_0`		0 dB	310 ms
`Reverb_-6`		−6 dB	500 ms
`Reverb_-8`		−8 dB	880ms
`t-rev`	time reversed speech	∞	0 ms
`pink`	pink noise	N/A	N/A
`mix`	time reversed speech + pink noise	N/A	N/A

Table 2: Parameters used for the ISM

Room dimension (m)	(8.0, 6.0, 3.0)
Source position (m)	(1.5, 1.0, 1.5)
Receiver position (m)	(7.0, 5.5, 1.5)

study investigates the effect of adding reverberation to a masking sound, ideally no reverberation originating from the acoustics of the testing environment should be added to the sound heard by the listeners. An anechoic chamber would meet this requirement. However, some feedback from preliminary testing conducted in an anechoic chamber pointed the unusual environment of an anechoic chamber, affecting listeners psychologically in terms of distraction and annoyance. Thus attempt was made to further reduce the reverberation in the listening room by introducing soft foams being attached to the walls of the listening room. The final reverberation time of this room was thus, measured to be approximately 0.3 s and the DRR measured at the listner's position with respect to the position of the target speech was about 7.3 dB. The level of ambient noise in the listening room was 18 dB(A).

3.2.2. Device setup

The device setup for this study is analogous to the configuration that has been used to conduct a similar test for sound masking system [24], as shown in Figure 4. The idea of this configuration is to simulate a working environment such as that of an open plan office.

Two loudspeakers were located in front of the listener's seat, one being 2.5 m away for projecting the masking sound (front loudspeaker) and another being 3.5 m away for projecting the target speech (back loudspeaker) to the listener. The height of the back loudspeaker was higher than that of the front loudspeaker to avoid the front loudspeaker blocking the path of sound

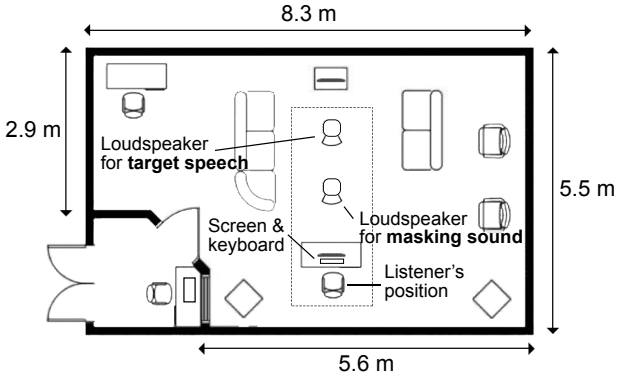


Figure 3: Configuration of the listening room.

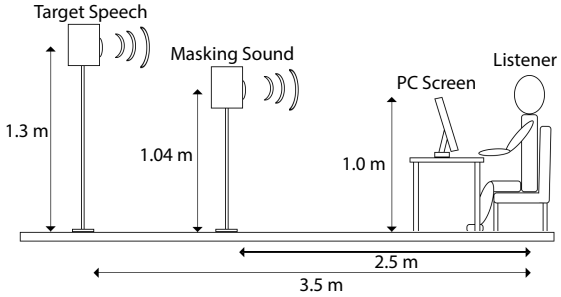


Figure 4: Position of loudspeakers and listener.

propagating from the back loudspeaker to the listener.

It is important that both loudspeakers were calibrated so that the target-to-masker ratio (TMR) was set to 0 dB, i.e. both loudspeakers to play a normalised sound at the same volume. To calibrate the loudspeakers, the volume of the back loudspeaker was initially adjusted to project a normalised speech file at the level of 58 dB(A) at 1.0 m away, i.e. front loudspeaker's position, since standard average speech has been scientifically proven to range between 55 to 60 dB(A) at a distance of 1.0 m [25]. The sound level at the listener's seat is then measured which was 52 dB(A). To ensure the TMR of 0 dB, the volume of the front loudspeaker was hence calibrated to have the same sound level of 52 dB(A) at the listener's seat.

As shown in Figures 3 and 4, a computer screen, a wireless keyboard, and a track pad were placed on a desk in front of the listener's seat. These devices were used for the listeners to enter their answers for the tests through graphical user interfaces (GUI) on the screen an example of which is shown in Fig. 5. The computer to which these devices were connected was placed outside the listening room to avoid the noise generated by the computer to be emanated into the listening room. A stereo sound file, which contains the normalised mask-

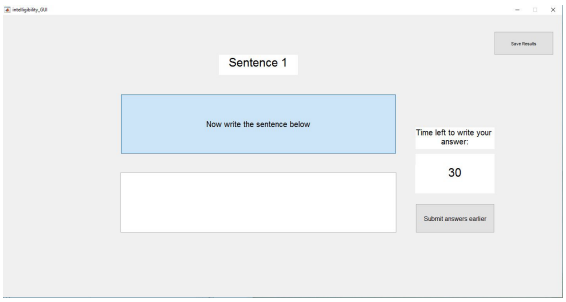


Figure 5: Example screen of the GUI used in the intelligibility measurement test.

ing sound and target speech on a different track, was played by this computer and was controlled by the GUI.

3.2.3. Participants

A group of 18 adult participants were tested, ranging from 20 to 25 years old and consisted of 13 males and 5 females. The participants were recruited from students at the University of Auckland, New Zealand. Based on their self-report, 9 were native and another 9 were bilingual speakers of English; with all of them having a normal hearing ability. The same listeners were tested throughout the three tests stated in Section 3.

3.3. Intelligibility measurement test

The testing method for measuring the intelligibility was developed based on the method used in [10], where the listeners were required to write down the whole sentence of the target speech while the masking sound was also played. The order of masked sentences and the speech masking algorithms to be used in the test was the same for all listeners.

To conduct the intelligibility test, a GUI was created, to allow the control of presenting the masked sentences and also to take in the answers the listeners provided afterwards. When the start button on the GUI was pressed, a target speech and masking sound were played simultaneously. Afterwards the listener had 30 s to type in their answer on a blank space provided on the GUI. After submitting their answer, the listener could press a submit button before the 30 s elapsed. To move on to the next sentence the listener pressed the start button.

A total of 21 Harvard Sentences were used to test for speech intelligibility level, with three different Harvard Sentences used for each of the seven speech masking sound. A list of Harvard Sentences used in the test is included in the Appendix. To process the intelligibility score for each Harvard Sentence with its respective masking sound, each correct word in a sentence was given one mark. Should the word attempted by listeners be of a similar phoneme to that of the correct word, e.g. “wide” and “white”, half a mark was given instead. The total marks attained were then compared against the total number of words in the respective Harvard Sentence to obtain its final score in percentage. Each Harvard Sentence contained a range of about 7 to 10 words, an average of 8 words in most cases. A high score in this case indicates a high intelligibility level i.e. speech masking sound is less effective in keeping speech privacy.

The listeners were allowed to try 3 example sentences beforehand, different to the sentences used in the formal test, to familiarise themselves with the GUI and the testing conditions.

3.4. Distraction measurement test

A cognitive test was conducted to measure the degree of distraction caused by masking sound. For this study, a memory test was selected to measure the distraction level because short-term memory is a simple process that almost anyone can perform without problems, and can easily change or fix the difficulty of the test by increasing the number of elements to memorise. The memory test has also been used in previous studies to measure the distraction level [14, 26], although not exactly the same as the one selected in this study.

The memory test presents the listener with 9 single digit integers ranging between 1 to 9 for 15 s, that they have to memorise in correct order, and type out the answer within 30 s. The 9 numbers are all randomly generated by a computer program, but the questions are consistent for all listeners. The numbers were presented in their numeric formats, not in words. In this test only the masking sound was presented thus the listeners evaluated the distraction caused by the masking sound itself rather than speech intelligibility as with the previous test.

Since it was found that short-term memory of humans was between 5 and 9 elements (averaging at 7 elements) [27], it was decided that the focus was intended towards the difficult-end of 9 elements. The choice of 15 s was decided by conducting a series of preliminary tests without masking sounds, to determine the absolute minimum time on average that listeners require to memorise all 9 numbers correctly. This means that if the masking sound were to have an influence in the listener’s concentration, it should lower their scores for the memory test, allowing a good comparison of which masking sound is the least distracting.

Each speech masking sound was tested with 3 memory questions, in total 21 memory questions for each listener. Each listener's score was marked out of 9, with one mark given for each correct number, and the percentage of correct answers was evaluated.

Although the algorithm for masking sound design (described in Section 2) remains the same, the underlying Harvard Sentences used to create the masking sounds are different from the ones used in the intelligibility test. The length of the masking sound for this test should be at least 15 s long in total whereas the masking sounds used in the intelligibility test were only 3 to 4 s long, which was the same length as the original Harvard Sentences used as the target speech. Based on feedback from preliminary tests it would be quite distracting and annoying if the same masking sound repeats, therefore several Harvard Sentences were concatenated to create a 15 s long masking sound to be used in the memory test.

Due to the fact that no target speech was played in this test, the back loudspeaker was not used, while the front loudspeaker played the masking sounds. The order of the masking sound presented in this test was randomised but the order was kept consistent for all listeners.

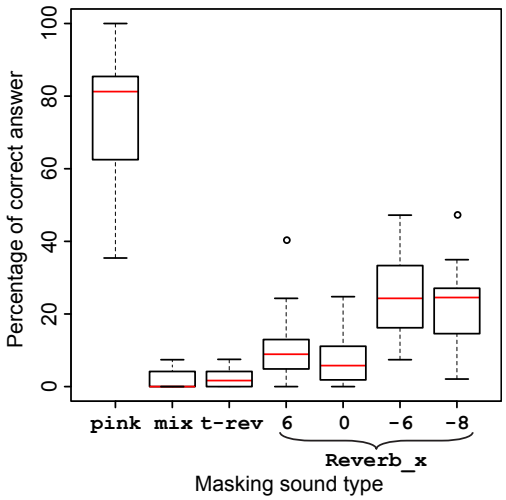
3.5. Annoyance measurement test

The final test of this study measured the listener's opinions about the level of annoyance caused by each masking sound through the use of the visual analogue scale (VAS) [28].

For the purposes of this research, the term *annoyance* was defined as "the varying degree of opinions on the level of discomfort the listeners are experiencing while not performing work". To summarise, the "distraction level" is different to the "annoyance level" where it is based on the listeners' personal opinion on the masking sounds when they are not working, but in a more relaxed state, whereas the "distraction level" measures how the masking sounds affect the listeners' concentration when they are working.

Listeners were required to evaluate seven different masking sounds by giving a score out of 100, with 100 being the "least annoying" and 0 being the "most annoying". For each masking sound, there were three sets for the listener to evaluate, meaning there were 21 files of masking sound to rate in total. The order of masking sound presented was randomised for each test set. Within the seven different masking sounds the listeners were allowed to listen to them multiple times.

In order to minimise skewing of data, the listeners were carefully advised to select the least annoying sound and give it a score of 100, then to select the most annoying sound and give it a score of 0, with the five remaining sound files to be rated against their least annoying sound. In essence, this provides data about the listener's opinion as to which sound was the least annoying and by how much, out of the seven sounds provided to them.



(a) Boxplot showing the distributions of intelligibility scores.

		pink	mix	t-rev	Reverb_x			
					6	0	-6	-8
pink			**0.019**	**0.020**	**0.020**	**0.020**	**0.020**	**0.020**
mix				53.662	**0.044**	**0.061**	**0.019**	**0.023**
t-rev					**0.159**	1.738	**0.020**	**0.020**
Reverb_x	6					17.001	0.245	0.327
	0						**0.020**	**0.137**
	-6							28.683

(b) p-values (%) of the Wilcoxon signed rank test for intelligibility scores. Values with red boldface show the pairs of masking sounds the null hypothesis of which is rejected by 5% significance level after applying the Bonferroni correction.

Figure 6: Results of intelligibility measurment test.

4. Results and discussion

4.1. Intelligibility measurement test

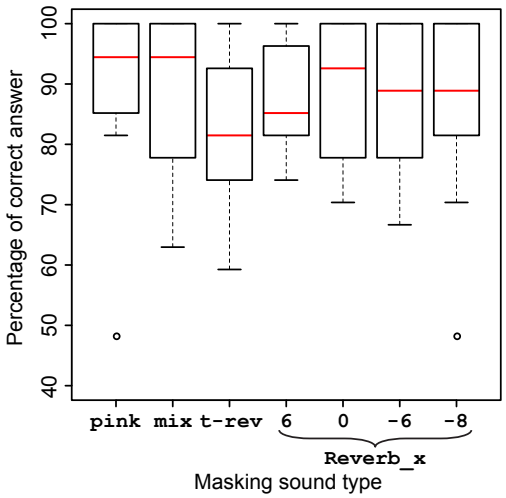
Figure 6 shows the distribution of the scores for the intelligibility test represented by a box plot. In the box plot red lines show the median of overall scores while the bottom and top edges of boxes are the first and third

quartiles, respectively. The whiskers show the 1.5 interquartile range. `mix` and `t-rev` showed the highest effectiveness while the `pink` was the least effective masking sound in terms of reducing the intelligibility of the target speech, which agrees with the results reported in previous studies [4].

For the proposed reverberant masking sound, a general trend can be observed that the intelligibility of the target speech is slightly improved (i.e. the masking effect was slightly reduced) by increasing the reverberation intensity. Although increasing reverberation intensity reduces the effectiveness of the masking effect, this does not indicate that the masking sound has lost its speech masking ability. The intelligibility scores for the reverberant masking sound with high reverberation intensity are still typically low (median of approximately 20%) compared to `pink` with much higher average scores (median of approximately 80%). By comparing the scores between the reverberant masking sound with different intensity of reverberation added, there is a relatively large gap between `Reverb_0` and `Reverb_-6` whereas the difference is marginal between other pairs.

Figure 6(b) shows the p-values of the Wilcoxon signed rank test [29] conducted to compare the differences of the medians across the seven different masking sounds. With the significance level $\alpha = 0.05$ (5 %) after applying the Bonferroni correction [30], it can be said that `pink` shows a significantly high intelligibility score compared to the other masking sounds all of which are generated from a speech-like masking sound. This indicates that pink noise, which is still widely used commercially, is less effective for reducing intelligibility of target speech. On the other hand, the result also shows significant differences between `t-rev` and the proposed reverberant masking sound with various DRRs except `Reverb_0`. This means only `Reverb_0` among the proposed masking sound with various DRRs achieves the effect close to that realised by `t-rev`. It should also be remarked that only `Reverb_0` shows a different trend compared to the reverberant masking sound with the other DRRs; there is no significant differences among `Reverb_6`, `Reverb_-6` and `Reverb_-8` whereas `Reverb_0` shows a significant difference from those other reverberant masking sounds. These facts, in conjunction with the findings from Fig. 6, indicate that among the proposed reverberant masking sound with various DRRs, `Reverb_0` seems to be the most effective masking sound to reduce intelligibility.

In summary the proposed reverberant masking sound is still quite effective in masking speeches, keeping a low speech intelligibility level.



(a) Boxplot showing the distribution of distraction scores.

		pink	mix	t-rev	Reverb_x			
					6	0	-6	-8
pink			55.095	0.911	15.238	29.921	3.946	17.792
mix				2.224	55.036	90.812	24.620	50.667
t-rev					8.516	1.224	24.705	12.443
Reverb_x	6					48.109	55.337	81.089
	0						25.114	44.397
	-6							91.286

(b) p-values (%) of the Wilcoxon signed rank test for distraction scores. None of the pairs of masking sounds, the null hypothesis of which is rejected by 5% significance level after applying the Bonferroni correction, is observed.

Figure 7: Results of distraction measurement test.

4.2. Distraction measurement test

Figure 7 shows the percentage of correct answers out of 9 numbers from the memory test, which means a high score indicates that the masking sound is less distracting.

Referring to Figure 7(a), it shows that the medians across all masking sounds ranges between 80% and 95% of correct numbers out of 9, or equivalently, listeners have medians ranging from 7 to 9 correct numbers across all masking sounds. The high medians across all listeners indicate that the memory test may have been too simple, considering that an average person's short-term memory already falls within the range of 7 ± 2 elements [27]. This experimental method suggests that the masking sounds would have to be exceptionally dis-

tracting to cause the listeners to only remember significantly less numbers outside of the short-term memory range.

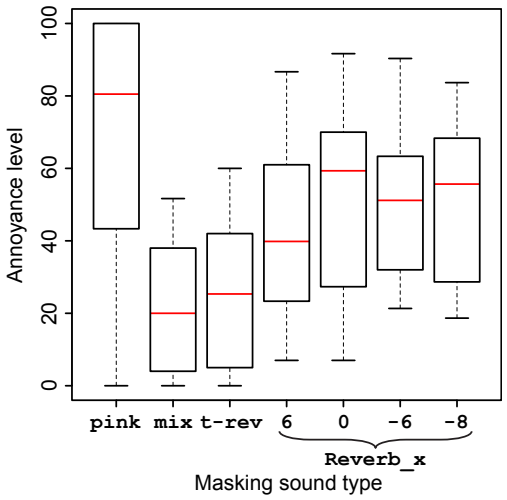
Figure 7(b) shows the p-values of the Wilcoxon signed rank test conducted to compare the differences of the medians across the seven different masking sounds. With the significance level $\alpha = 0.05$ (5 %) after applying the Bonferroni correction, the result indicates that no specific masking sounds are significantly distracting the listeners from memorising 9 numbers in order. The results for the memory test does not provide statistically sufficient proof which masking sounds distract the listeners more. This would have been influenced by the design of the memory test that requires the memorisation of only 9 numbers. However, in fact, a trend can be still seen across the masking sounds where `pink`, `mix`, and `Reverb_0` have the highest medians verses the most distracting sound of `t-rev`. It is believed that a better design of a distraction test with a larger scale of results can provide better insight and proof into the reliability of this trend.

4.3. Annoyance measurement test

For the annoyance test the scores across three masking sounds generated from different sentences are averaged for each type of masking sound, then analysed statistically across all listeners. The results from this analysis is provided in Figure 8, with the scores out of 100 being the least annoying sound based on the listeners' opinions.

Figure 8(a) shows the median of the reverberant sounds to be higher than that of `mix` and `t-rev`, while still lower than `pink`. However it can also be seen that `pink` appears to have mixed results as some listeners believed it to be the most annoying sound in some cases.

Figure 8(b) shows p-values of the Wilcoxon signed rank test conducted to compare the differences of the medians across the seven different masking sounds. In the result, with the significance level $\alpha = 0.05$ (5 %) after applying the Bonferroni correction, `pink` is significantly less annoying compared to `mix`, `t-rev` and `Reverb_6`. The proposed reverberant masking sounds are also significantly less annoying than the `mix` and `t-rev` regardless of the intensity of reverberation added, indicating that addition of reverberation proves effective in reducing the annoyance. Conversely, among the proposed masking sounds, there is no significant difference from each other. Overall these observations show that the proposed reverberant masking sounds can reduce the level of annoyance, which may reach the level close to that of the pink noise that is providing minimal annoyance.



(a) Boxplot showing the distribution of annoyance scores.

		pink	mix	t-rev	Reverb_x			
					6	0	-6	-8
pink			**0.000**	**0.000**	**0.069**	2.748	2.232	5.411
mix				40.277	**0.176**	**0.002**	**0.002**	**0.000**
t-rev					**0.006**	**0.000**	**0.000**	**0.003**
Reverb_x	6					1.180	8.163	6.488
	0						79.391	82.954
	-6							66.514

(b) p-values (%) of the Wilcoxon signed rank test for annoyance scores. Values with red boldface show the pairs of masking sounds the null hypothesis of which is rejected by 5% significance level after applying the Bonferroni correction.

Figure 8: Results of annoyance measurement test.

4.4. Most effective masking sound

By looking into the results across the three listening tests a few conclusions can be drawn.

Firstly, the results clearly show that adding reverberation to speech-like masking sound has an effect to mitigate the annoyance of the masking sound while maintaining its masking effect at sufficient level. This supports the hypothesis that reverberation plays a role as a filter that removes the pulsive components in the waveform of conventional speech-like masking sound causing annoyance while the major spectral components of the masking sound that contribute to mask target speech are kept.

Secondly, by comparing different intensities of reverberation added to the masking sound, of those several

DRRs chosen in the listening test, the DRR of 0 dB is the best, compromising the trade-off between the masking effect (intelligibility) and detrimental psychological effect (annoyance). In other words, the appropriate intensity of reverberation added to get the most effective performance would be the same energy as that of the original masking sound.

Finally, the distraction caused by masking sound may be slightly eased by adding reverberation to the masking sound. Unfortunately, the distraction test designed in this study could not show a statistically significant difference in the level of distraction between different masking sound. However, by looking at the trend observed in the study, it is still likely that reverberation would have a positive effect to reduce the distraction level. Further studies need to be conducted by using different testing methodology to measure the distraction for proving this.

5. Conclusion

In this study, effect of adding reverberation to masking sound has been investigated. The study stood on a hypothesis that adding reverberation to a speech-like masking sound would mitigate the distraction and annoyance caused by the masking sound while keeping its effectiveness to reduce the intelligibility of a target speech to be masked. Artificial room impulse responses with different intensity of reverberation were simulated by the image source method which was then convolved with a time-reversed speech to generate a reverberant speech-like masking sound. The performance of the designed masking sound was tested by subjective listening tests examining the suggested three different aspects, namely intelligibility, distraction and annoyance. The time-reversed speech, which has been reported as the most effective masking sound to reduce the intelligibility of target speech, was selected as the seed of the proposed masking sound. The experimental results proved that adding reverberation had an effect to reduce the annoyance caused by the masking sound while maintaining its performance to hide the information in target speech.

A further study is required to investigate the effect of different TMR. It would also be worthwhile to study the effect of reverberation using room impulse responses measured in an actual reverberant room. The findings from this study also imply that the reverberation has some effect to mitigate the annoyance caused by speech-like masking sound. Since reverberation can be deemed as a filter from signal processing point of view, more detailed study on pursuing the optimal design of filter that provides the more effective but less annoying and distracting masking sound would be needed.

Acknowledgement

The authors would like to thank all anonymous participants of the subjective listening tests conducted for this study. Many thanks to Dr Catherine Watson at the University of Auckland for her comments on the statistical analysis of the experimental results. This work was supported by Faculty Research Development Fund at the University of Auckland.

Appendix A. List of Harvard Sentences

In the listening test following sentences included in the Harvard sentences were used for the target speech as well as the seed speech to generate the masking sound.

1. The sky that morning was clear and bright blue.
2. The boss ran the show with a watchful eye.
3. Take the winding path to reach the lake.
4. The streets are narrow and full of sharp turns.
5. The idea is to sew both edges straight.
6. They sang the same tune at each party.
7. The shelves were bare of both jam or crackers.
8. Crouch before you jump or miss the mark.
9. The fish twisted and turned on the bent hook.
10. A smatter of French is worse than none.
11. Either mud or dust are found at all times.
12. Torn scraps littered the stone floor.
13. The sense of smell is better than that of touch.
14. It was a bad error on the part of the new judge.
15. A six comes up more often than a ten.
16. The couch cover and hall drapes were blue.
17. Drop the ashes on the worn old rug.
18. The beam dropped down on the workmen's head.
19. It matters not if he reads these words or those.
20. What joy there is in living.
21. Watch the log float in the wide river.

References

[1] J. Bradley, Acoustical design for open-plan offices, Construction Technology Update 63 (2004) 1–6.

[2] S. Stapels, Human response to environmental noise: Psychological research and public, American Psychologist 51 (2) (1996) 143–150.

[3] N. Clark, D. Rose, Y. Hioka, Effect of using artificial echoes for keeping speech privacy, in: 12th Western Pacific Acoustics Conference 2015, 2015.

[4] A. Ito, A. Miki, Y. Shimizu, K. Ueno, H. Lee, S. Sakamoto, Oral information masking considering room environmental condition part1: Synthesis of maskers and examination on their masking efficiency part1 synthesis of maskers and examination of their masking efficiency, in: Proceedings of Inter-Noise 2007, 2007.

[5] K. Ueno, H. Lee, S. Sakamoto, A. Ito, M. Fujiwara, Y. Shimizu, Oral information masking considering room environmental condition Part1: Synthesis of maskers and examination on their masking efficiency part2: Subjective assessment for "masking efficiency and annoyance", in: Proceedings of Inter-Noise 2007, 2007.

[6] K. Ueno, H. Lee, S. Sakamoto, A. Ito, M. Fujiwara, Y. Shimizu, Experimental study on applicability of sound masking system in medical examination room, in: Acoustics08 Paris07, 2008, pp. 2374–2378.

[7] H. Lee, K. Ueno, S. Sakamoto, M. Fujiwara, S. Shimizu, M. Hata, Effect of room acoustic conditions on masking efficiency, in: Proceedings of Inter-Noise 2009, 2009.

[8] Y. Hara, K. Miyoshi, Y. Shimizu, M. Fujiwara, Estimation of masking effects on speech according to spectral and dynamical characteristics of maskers, in: Proceedings of Inter-Noise 2009, 2009.

[9] J. Keraenen, V. Hongisto, Achieving speech privacy with reasonable sound insulation and masking background noise, in: Proceedings of Inter-Noise 2009, 2009.

[10] M. Fujiwara, Y. Shimizu, M. Hata, H. Lee, K. Ueno, S. Sakamoto, Experimental study for speech privacy with a sound masking system in a medical examination room, in: Proceedings of Inter-Noise 2009, 2009.

[11] A. G. DeLoach, J. P. Carter, J. Braasch, Tuning the cognitive environment: Sound masking with natural sounds in open-plan offices, The Journal of the Acoustical Society of America 137 (4) (2015) 2291–2291.

[12] M. Leek, M. Brown, M. Dorman, Informational masking and auditory attention, Perception & Psychophysics 50 (3) (1991) 205–214.

[13] T. Arai, Masking speech with its time-reversed signal, Acoustic Science & Technology 31 (2) (2010) 188–190.

[14] B. Jing, A. Liebl, P. Leistner, J. Yang, Sound masking performance of time-reversed masker processed from the target speech, Acta Acustica United with Acustica 98 (4) (2012) 135–141.

[15] A. Nábělek, P. Robinson, Monaural and binaural speech perception in reverberation for listeners of various ages, The Journal of the Acoustical Society of America 71 (5) (1982) 1242–1248.

[16] A. Nábělek, T. Letowski, F. Tucker, Reverberant overlap- and self-masking in consonant identification, The Journal of the Acoustical Society of America 86 (4) (1989) 1259–1265.

[17] J. Allen, D. Berkley, Image method for efficiently simulating small-room acoustics, The Journal of the Acoustical Society of America 65 (4) (1979) 943–950.

[18] E. A. Lehmann, A. M. Johansson, Prediction of energy decay in room impulse responses simulated with an image-source model, The Journal of the Acoustical Society of America 124 (1) (2008) 269–277.

[19] Y. Hioka, K. Niwa, S. Sakauchi, K. Furuya, Y. Haneda, Estimating direct-to-reverberant energy ratio using D/R spatial correlation matrix model, IEEE Transactions on Audio, Speech, and Language Processing 19 (8) (2011) 2374–2384.

[20] S. Mosayyebpour, H. Sheikhzadeh, T. A. Gulliver, M. Esmaeili, Single-microphone lp residual skewness-based inverse filtering of the room impulse response, IEEE Transactions on Audio, Speech, and Language Processing 20 (5) (2012) 1617–1632.

[21] IEEE recommended practices for speech quality measurements, IEEE Transactions on Audio and Electroacoustics 7 (3) (1969) 225–246.

[22] H. Kuttruff, Room Acoustics, 5th Edition, Applied Science Publishers LTD, 1973, Ch. 2.

[23] G. Schmid, A. Chan, Development and commissioning of an IEC standard listening room and two research applications, in: Proceedings of Bienniel Conference of the New Zealand Acoustical Society, 2008.

[24] T. Sannohe, T. Arai, K. Yasu, A method of creating speech masker using a database in a sound masking system, The Journal of the Acoustical Society of Japan 71 (2015) 382–389, (in Japanese).

[25] W. O. Olsen, Average speech levels and spectra in various speaking/listening conditions, American Journal of Audiology 7 (1) (1998) 21–25.

[26] C. Beaman, N. Holt, Reverberant auditory environments: The effects of multiple echoes on distraction by 'irrelevant' speech, Applied Cognitive Psychology 21 (8) (2007) 1077–1090.

[27] G. Miller, The magical number seven, plus or minus two: Some limits on our capacity for processing information, Psychological Review 63 (2) (1956) 81–97.

[28] I. Adamchic, B. Langguth, C. Hauptmann, P. Tass, Psychometric evaluation of visual analog scale for the assessment of chronic tinnitus, American Journal of Audiology 21 (2) (2012) 215–225.

[29] M. Hollander, D. A. Wolfe, E. Chicken, The One-Sample Location Problem, John Wiley & Sons, Inc., 2015, pp. 39–114. doi:10.1002/9781119196037.ch3.

[30] F. Curtin, P. Schulz, Multiple correlations and bonferronis correction, Biological Psychiatry 44 (8) (1998) 775 – 777. doi:http://dx.doi.org/10.1016/S0006-3223(98)00043-2.