Applied Acoustics 114 (2016) 171–178

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

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

ARTICLE INFO

Article history:
Received 1 May 2016
Received in revised form 4 July 2016
Accepted 20 July 2016

Keywords:
Speech privacy
Masking
Room reverberation
Intelligibility
Distraction
Annoyance

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.

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 constraints 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 [3–10]. 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 used 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 efficiency. 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

* Corresponding author.
E-mail address: yusuke.hioka@ieee.org (Y. Hioka).

http://dx.doi.org/10.1016/j.apacoust.2016.07.014
0003-682X/

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 listening 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

Fig. 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 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 and Johansson [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

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

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.

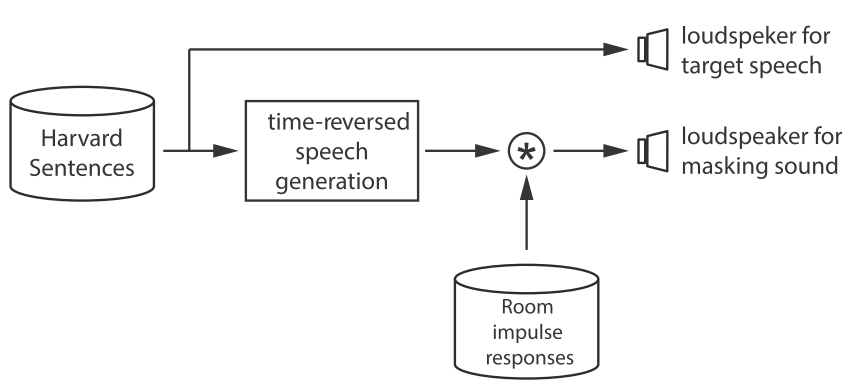

Fig. 1. Masking sound generation process.