Applied Acoustics 137 (2018) 151–160

Contents lists available at ScienceDirect

Applied Acoustics

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

Auditory distraction by speech: Can a babble masker restore working memory performance and subjective perception to baseline?

Tobias Renz[a,*], Philip Leistner[a,b], Andreas Liebl[b]

[a] Institute for Acoustics and Building Physics, University of Stuttgart, Pfaffenwaldring 7, 70569 Stuttgart, Germany
[b] Fraunhofer Institute for Building Physics, Nobelstrasse 12, 70569 Stuttgart, Germany

ARTICLE INFO

Keywords:
Sound masking
Background speech
Open-plan office
Speech privacy
Performance
Annoyance

ABSTRACT

Background speech has a detrimental impact on employees' privacy perception and cognitive performance in open-plan offices. Sound masking that covers speech sounds can help to improve the speech privacy and the ability to work undisturbed in open office environments. Recently, non-artificial sounds such as babble or water sounds have been suggested as masking sounds because they may be perceived as more natural and subsequently less annoying. This paper compares the working memory performance and annoyance perception during background speech that was masked by different babble sounds and a waterfall sound to speech that was masked by stationary noise with the frequency spectrum of the distracting background speech signal (target speech). In a first laboratory experiment the effects of different babble sounds were compared with the noise sound at a speech-to-noise ratio of −3 dB. All subjects had to complete a serial short-term memory task and a questionnaire that covered subjective ratings. All sound conditions with masked target speech produced similar error rates and annoyance ratings as unmasked target speech. In a second laboratory experiment, the speech-to-noise ratio of target speech that was masked by babble consisting of 48 target speech signals was varied between −6 and −12 dB in steps of 3 dB. Target based babble resulted in similar error rates as stationary noise at −6 dB speech-to-noise ratio but the findings suggest that target speech that is masked by babble might even be perceived as more annoying than speech that is masked by stationary target spectrum based noise.

1. Introduction

The quality of open-plan offices depends highly on the acoustic conditions. Acoustic privacy is the least satisfying aspect in office buildings according to a survey of 45,464–52,138 building occupants [1]. Disturbing background speech has a detrimental impact on the well-being of employees and their ability to concentrate at work. Current standards for open-plan offices, such as ISO 3382-3:2012 [2] or VDI 2569:2016 [3], focus on the reduction of the disturbing effects of background speech sounds. Sound absorbers and screens can reduce the sound pressure level (SPL) of disturbing background speech. However, a certain background noise level is necessary to cover these sounds.

The use of a sound masking system creates a controlled background noise. It is commonly accepted that all three means, sound absorbers, screens, and masking, should be used at the same time to condition an open office environment (e.g. [4,5]). The application of different masking sounds shows different effects on the working memory performance and annoyance perception of persons that are subjected to partially masked speech sounds. Recently, babble sounds have received attention in the context of sound masking and have been pointed out as favourable as compared to conventional masking sounds like stationary noise. This paper provides first a literature review of past studies on the effects of babble as a sound masking signal. This is followed by reporting two laboratory experiments that tested the beneficial effects of sound masking with babble on serial recall performance and annoyance perception. In conclusion, the potentials and limitations of different masking sounds are outlined.

1.1. Background

Background noise in open-plan offices is often characterised by speech sounds with changing-state features that affect the short-term memory [6,7]. Short-term memory for order of visually or acoustically presented items is impaired when sounds with temporal-spectral variability are presented simultaneously [7,8]. Colle and Welsh [9] were the first who reported that working memory performance is impaired when it is exposed to background sounds with sufficient variation in time and frequency. This effect is named Irrelevant Sound Effect

* Corresponding author.
E-mail address: tobias.renz@iabp.uni-stuttgart.de (T. Renz).

https://doi.org/10.1016/j.apacoust.2018.02.023
Received 23 November 2017; Received in revised form 23 February 2018; Accepted 26 February 2018
Available online 30 March 2018

[10] (for an overview see e.g. [11,12]). Semanticity can deteriorate the working memory performance additionally, for instance, when intelligible speech sounds are presented [13–15]. Serial recall is an established task to test working memory performance. Irrelevant background speech can also impair more complex tasks such as proofreading [16–18], reading comprehension [19,20], writing [21,22], and mental arithmetic [7,23]. Adding a masking sound to background speech can restore the working memory performance (e.g. [13,24,25]). There are two models to predict the decrease of working memory performance due to background sound. The first model makes use of the Speech Transmission Index (STI) and the second model uses Fluctuation Strength (FS) as predictor [26,27].

The open-plan office standard ISO 3382-3:2012 [2] for room acoustic measurements suggests parameters that are based on one simulated speech source present at a time. A recent study presented by Yadav et al. [28] concludes that multi-talker environments (e.g. when 4 talkers are simulated simultaneously) are a more realistic assumption, which leads to a higher performance impairment and is perceived as more distracting. The new draft of the German standard VDI 2569:2016 [3] refers to a natural masking effect of babble that can occur when multiple persons are talking at the same time. According to the results of laboratory experiments, the occurrence of a beneficial impact of natural babble on serial recall performance is improbable when the background talkers and the disturbing speech source are spatially separated [28,29]. When 4 employees at different workplaces speak at the same time, they may be even more disruptive than 1 talker present [28]. Jones and Macken [30] were one of the first who analysed the effect on working memory when the number of background voices is increased. Whilst 1 and 2 voices resulted in comparable performance decrements, monaurally presented babble of 6 voices reduced the performance impairment significantly. However, when these 6 voices were simulated at different spatial locations, the disruptive capacity was restored. Babble with 100 voices produced significantly fewer errors than a single voice, but significant disruption as compared to a silent sound condition [31,32].

Besides the occurrence of natural babble within an open space office, babble sounds have been tested as a sound masking signal for disruptive speech [24,33,34]. Babble sounds consisting of 8–128 voices lead at a signal-to-noise ratio (SNR) of −6 dB to similar consonant identification scores [35]. In the following, the SNR refers to the ratio of the A-weighted SPL of target speech to the A-weighted SPL of a masking sound. The speech intelligibility increases below 8 and above 128 voices. Hence, babble with 8–128 voices is a more effective target speech masker than stationary speech-shaped noise with regard to speech intelligibility. The lower speech intelligibility of speech masked by babble instead of speech-shaped noise may also result in higher working memory capacities. Therefore, it may be beneficial to use a babble sound instead of stationary noise that is commonly used in practical applications because babble decreases the speech intelligibility at higher SNRs and may sound more natural in an office environment. Brocolini et al. [33] tested the effect of speech masked by babble with 1, 3 and 5 voices on working memory performance and compared it to speech that was masked by a stationary noise. The SNR was set to −4 dB in all sound conditions, but the effect of unmasked target speech was not tested. All sound conditions led to a decrease in performance as compared to a control condition with stationary noise but the type of masking noise did not show a significant effect on the decrease of performance. The mentioned study of Jahncke et al. [34] suggests that the use of babble with 7 voices as a masking sound that is played over headphones is not as effective as the use of a nature sound consisting of bird twitter and rippling water. The serial recall performance was significantly higher when target speech was masked by the nature sound than during a sound condition with unmasked target speech. However, using babble as masking sound resulted in performance decrements that were similar to unmasked target speech. Keus van de Poll et al. [24] used a recording of 9 different voices, superposed it four times to the original recording and added a time shift between each of these 5 tracks. Hence, the babble consisted of 45 speech signals. As compared to a stationary noise sound with a decline of 5 dB per octave, the babble sound proved to be more suitable for masking target speech because the performance in serial recall was higher and the sound environment with this masking sound was perceived as more pleasant. The use of water waves as masking sound led to similar results as the babble sound, the mean serial recall scores were similar to silence and the sound environment was perceived as more pleasant than unmasked target speech. Based on these findings, monaural presentation of babble is a suitable sound masking signal for disturbing background speech as long as the number of voices is high, for instance, when the babble sound consists of 45 speech signals and the SNR is set appropriately (e.g. −3 dB as chosen by Keus van de Poll et al. [24]). It is unclear if babble sounds with less than 8 voices are suitable masking sounds. The temporal-spectral variability of a background sound is known to have a high impact on working memory performance and decreases with increasing number of voices. One may assume that 8 or more speech signals are needed to create babble that masks background speech beneficially with regard to working memory performance because speech intelligibility and FS are higher for less than 8 voices.

1.2. Aim of this study

While the speech intelligibility of speech masked by babble has been determined for multiple SNRs, number of voices, and types of voices, like the target speech signal as well as voices of same and different sex (e.g. [36]), the working memory performance has only been tested for a few specific configurations. Previous studies on the effects of babble as a masking sound on the working memory performance considered only one SNR and did not vary the number of voices of the babble sound between 8 and 128 voices. This study addressed if sound conditions with non-artificial masking sounds, namely a waterfall sound and different babble sounds, have a beneficial impact on working memory performance and are perceived as less annoying than artificial masking sounds.

Two laboratory experiments were performed that tested the effect of different babble masking sounds on working memory performance and perceived annoyance. In Experiment 1 different masking sounds were superimposed to a speech recording (target speech) at −3 dB SNR while the SNR was lowered to a range between −6 and −12 dB in Experiment 2. Both experiments included 2 control conditions, silence and unmasked target speech. Experiment 1 compared the effectiveness of different masking sounds such as noise-like masking sounds (stationary target based noise and noise with a decline of 5 dB per octave), nature sounds (waterfall and wind sounds), and speech-like masking sounds (time-reversed speech and babble). This study did not focus on the sound conditions with the time-reversed masker and the wind sound but they were added to test the effects of different masking signals. Five babble sounds with different voices were tested, 12 different male voices, 12 same male voices, 12 same female voices, 12 target voices, and 48 target voices, respectively. Experiment 2 compared babble with 48 target voices to babble with 48 target voices that was adjusted to a frequency spectrum that decreased by 5 dB per octave. Target based noise and the same waterfall sound as used in Experiment 1 were added to enable a comparison to non-fluctuating masking sounds.

STI and FS values of all sound conditions were calculated. According to both models, the predicted error rates for all sound conditions with masked target speech were within the error rates of the 2 control conditions, unmasked target speech and silence. FS of the sound condition with the babble of 48 voices is slightly smaller than of the babble with 12 voices, and hence target speech that is masked by one of these two masking sounds at negative SNRs results in similar FS values. According to Schlittmeier's model [27] only small improvements in serial recall performance were expected. The sound condition with