DOI: 10.61782/fa.2023.0366

EVALUATING ACOUSTIC ALARM AUDIBILITY IN NOISY ENVIRONMENTS: AN AUTOMATIC APPROACH USING DEEP LEARNING

François Effa[*1,2,3] **Jean-Pierre Arz**[1] **Nicolas Grimault**[2] **Romain Serizel**[3]

[1] Institut National de Recherche et de Sécurité, F-54000 Nancy, France

[2] Centre de Recherche en Neurosciences de Lyon, CNRS, F-69500 Bron, France

[3] Université de Lorraine, CNRS, Inria, Loria, F-54000 Nancy, France

ABSTRACT

In noisy occupational settings, the audibility of acoustic alarms is crucial to alert workers to potential dangers. To ensure the audibility of warning signals, the international standard ISO 7731 requires that alarms be "clearly audible" and provides the level to play the auditory alarms relative to ambient noise. However, the expression "clearly audible" is not defined and the level specifications lead to excessive alarm levels for high noise levels, thereby exposing workers to dangerous high sound levels. In order to ensure that alarms are audible without being excessively loud, the goal of our work is to propose an automatic approach to assess their audibility at supraliminal levels, thus avoiding having to test them experimentally. Our contribution is: (1) the development of an experimental method to assess the audibility of the alarms according to ISO 7731; (2) using this method, the collection of extensive perceptual data on a minimum of 12 normal-hearing listeners (minimum 2000 training samples involving 70 alarms and 50 noises); (3) the development of a convolutional neural network that is learned and evaluated from the perceptual data, showing strong generalization; (4) a detailed analysis of both the evaluation data and the model performance.

Keywords: *alarm, audibility, deep learning, occupational safety*

[*]***Corresponding author***: francois.effa@inrs.fr

1. INTRODUCTION

The ability to perceive and respond to auditory danger signals in work areas is paramount for ensuring workplace safety and preventing accidents. To ensure that acoustic alarms are easily heard and recognized in their intended environment, the ISO 7331 standard [1] requires alarms to be "clearly audible" and establishes a set of three criteria based on time-averaged technical measures.

It should be noted that, in the field of psychoacoustics, the expression "clearly audible" lacks a clear and precise meaning. According to the standard, an alarm is "clearly audible" if at least one of the three criteria is met. The first criterion, which requires a 15 dB-difference between the respective A-weighted sound pressure levels of the signal and the ambient noise, is the most frequently used because of its low complexity. However, it leads to excessive alarm levels, especially in very noisy work environments [2]. A 15 dB signal-to-noise ratio (SNR) between the alarm and the ambient noise may therefore be counterproductive from the point of view of occupational health and safety, since it can cause permanent damage to the workers' auditory system or startle reactions that put them at risk [1, 2].

In the context of occupational health and safety, it is imperative to find an optimal balance to ensure that alarms are audible enough to produce an adequate reaction without being excessively loud. Faced with the plurality of criteria proposed in the literature, Laroche *et al*. point out the lack of guidelines systematically used in the field to set the level of auditory alarms [3]. At this point, there are two solutions available. The first solution is to adjust alarm levels based on listening tests. Nevertheless, it is time-consuming and limited by its stimulus-dependent aspect, since the perception of an alarm in noise relies on spectro-temporal characteristics of both the signal and the noisy background [4]. Consequently, any change in the alarm or background noise requires new listening tests, making this approach unsuitable for an application to a wide range of signals. The second solution is to use predictive approaches such as existing auditory models. Several models have demonstrated good reliability in predicting masked thresholds of signals in complex acoustic environments [5, 7]. However, these models are limited to predicting detection thresholds and do not provide information about perception of target signals at higher levels, which is necessary for auditory alarms. Furthermore, using such models to adjust the level of acoustic alarms would require an explicit criterion for assessing whether an alarm is "clearly audible" or not. Yet, such criterion does not exist, as the expression "clearly audible" remains undefined.

In this study, we propose a deep learning-based approach to address the problem of characterizing alarm signals as "clearly audible" or not "clearly audible". As a data-driven paradigm, supervised learning can effectively accomplish this task by using knowledge acquired from training examples. Previous studies have shown the high accuracy of deep neural networks in audio classification tasks [8], indicating their proficiency at learning intricate patterns in audio data.

The present work has a dual purpose. First, through an experiment carried on normal-hearing subjects, it aims to clarify the term "clearly audible" expressed in the ISO 7731 standard as a requirement for acoustic alarms in work areas. Second, it introduces an automatic approach using a convolutional neural network (CNN) to predict the audibility of alarms in noise. In this paper, we report the method employed to experimentally assess the audibility of acoustic alarms in noise. We also describe the development of a CNN, trained and evaluated on the experimental data. The results will be presented at the conference.

2. METHODS

2.1 Experimental evaluation of audibility

Most of the time, the notion of audibility refers to the probability of detecting a sound under specific conditions. Thus, one may consider to study the question of the audibility of acoustic alarms in noise through masked thresholds. However, in order to effectively elicit a response, danger signals must be played at levels well above these masked thresholds. According to different studies, alarms should exceed the masked thresholds by 12 to 25 dB for optimal use [3, 9-11], but such a wide range is of limited interest for practical application. As a result, the masked threshold for a given alarm is not sufficient to establish a reliable audibility criterion. In order to thoroughly investigate the perception of audibility in relation to acoustic alarms and critically examine the terminology used in the standard, we designed an experiment that included both the measurement of masked threshold, and the assessment of the "clearly audible" attribute of auditory warning signals in noise. The results of this experiment will allow for an analysis of the relationships between detectability and perceived "audibility" of warning signals at higher levels, as well as a comparison with the recommendations of the standard.

2.1.1 Participants

Twenty volunteers between 20 and 50 years old took part to the experiment. They all had an average tone loss of less than 20 dB HL across the frequencies 500 – 1000 – 2000 – 4000 Hz. They received financial compensation for the time spent on the experiment.

2.1.2 Stimuli and material

The stimuli were 5.5-second mono sound clips, created by mixing an alarm with a field recording of a noisy workplace. The alarms and backgrounds were collected from multiple sources, either public (Freesound [12], BigSoundBank [13]) or personal recordings. The signals were split into five contextual categories: (1) construction noise and reverse alarms, (2) railway and construction noise, warning signals for track workers, (3) urban traffic and car horns, (4) factory noise, miscellaneous buzzers and indoor warning signals, (5) locomotive cab noise, warning signals for train drivers. Each category contained 2 backgrounds and 3 alarms. In total, we created 30 alarm-background pairs by associating the alarms and backgrounds within each category. Mixing was performed by adding the alarm to the noise with a pseudo-random temporal onset, carefully controlled to avoid onsets too close to the beginning or the end of the clip.

The experiment took place in a soundproof room. The stimuli presentation and participant responses were managed by a custom MATLAB application. The signals were processed through an RME Babyface Pro soundcard and presented over circumaural headphones (Beyerdynamic DT 770 Pro), calibrated with Larson Davis AEC101 artificial ear and Model 824 sonometer.

2.1.3 Procedure

The experiment was made of two tasks alternating in a random order: a detection task, and an audibility

assessment. In both tasks, the 30 alarm-background pairs were evaluated, and the stimuli were presented in a random order for each participant.
The detection task was based on a two-alternative forced choice paradigm and the method of constant stimuli. A trial consisted in a presentation of two consecutive intervals spaced with a 500 ms pause. The two intervals contained the same background noise, but only one of them contained an alarm. After the end of the second interval, the participants were asked to report in which of the two intervals they had heard the alarm. Six different SNRs, spanning from -30 to 0 dB, were used to present the stimuli. The experiment included two levels of noise, namely 60 and 80 dBA. For a given SNR and noise level, each clip was presented three times per listener.
For the audibility assessment, the participants were instructed to listen passively and not to focus on detecting alarms. This task followed a Yes-No design, using the method of constant stimuli. In a trial, the listeners were presented with a single clip containing the alarm-background mix. After the presentation, they had to respond with either "yes" or "no" to the question "*Was the alarm clearly audible?*". The stimuli were presented at six SNRs ranging from -22.5 to +10 dB, and two levels of noise: 60 and 80 dBA. As in the detection task, each condition was repeated three times per listener.

2.2 Audibility prediction using a CNN

To predict the audibility of acoustic alarms using deep learning, a model was trained to perform the same audibility assessment as in the human experiment. This amounts to define the problem as a binary classification task. Taking spectro-temporal representations of the 5.5-second sound clips as input, our model is designed to produce a binary prediction indicating whether the alarm in the clip is "clearly audible" or not.

2.2.1 Dataset

The dataset contains 5.5-second alarm-background mixes as described in Section 2.1. These signals come with perceptual annotations, that are the responses of the participants to the question "*Was the alarm clearly audible?*". The dataset is composed of two subsets that were not collected at the same time: development (training and validation) and evaluation.
To evaluate the model, we need reliable and interpretable data. Therefore, the data collected in the experiment from Section 2.1 will serve as evaluation data, since the annotation procedure is well controlled and follows the standards of usual psychoacoustical experiments (repeated measures design, fixed SNRs and noise levels, multiple repetitions per participant). As a result, the evaluation subset contains 360 sound clips (30 alarm-background pairs at 6 SNRs and in 2 levels of noise), each clip having 60 annotations.
In contrast, development subset does not require as much constraints as the evaluation subset. For development, we have 2000 clips, made with completely different alarms and backgrounds than the evaluation subset, all downloaded from the Freesound database. The alarm-background pairs were generated randomly among a total of 70 alarms and 52 backgrounds. Each clip was mixed at a random integer SNR between -30 and +15 dB. The noise level was randomly selected to be either 60 or 80 dBA. The clips were all annotated by 10 listeners, with 8 of these listeners providing annotations for both the development and evaluation subsets. To speed up the data collection, in the annotation procedure, each clip was presented only once per annotator instead of three times for the evaluation subset. A 20% of this data was randomly split for validation purposes.
For both development and evaluation, since all the clips were annotated by multiple listeners, the final labels for each clip (0: not clearly audible, 1: clearly audible) were obtained based on the majority responses.

2.2.2 System

The first stage of the system is feature extraction, that is, the computation of the spectro-temporal representations used by the model. For that purpose, the signals were sampled at a frequency of 44.1 kHz and mel-spectrograms with 64 coefficients were extracted as input features to the model. The extraction process involved a 1024-sample short-time Fourier transform (STFT) using a Hamming window with a 50% overlap.
The design of the model was guided by both preliminary experiments and a review of the existing literature [14-16]. The model is a neural network with 4 convolutional layers that have 32, 64, 64 and 128 filters, respectively, using a 3-by-3 kernel size. After every convolutional layer, ReLU activations are applied followed by max pooling along the frequency axis, with kernel sizes 4, 4, 2, and 2, respectively. The activations of the last convolutional layer are then stacked along frequency axis and passed through a time-distributed fully connected layer and sigmoid activation. The output representation, of size (128, 1), is finally subjected to L_p aggregation with parameter $p = 5$. Following aggregation, we obtain a value that ranges from 0 to 1, which tends to approach 1 when the alarm sound in the clip is clearly audible, and 0 when it is not.

3. CONCLUSION

The aim of this study is to provide a precise definition of the term "clearly audible" as a requirement for acoustic alarms in work areas and proposes an automatic approach using a convolutional neural network to predict the audibility of auditory danger signals in noisy work environments. An experiment was designed to evaluate the audibility of acoustic alarms. It encompasses the measurement of masked thresholds and the assessment of the "clearly audible" attribute of auditory danger signals in noise in order to establish a reliable audibility criterion. The proposed approach using a convolutional neural network has the potential to provide an automated solution for adjusting alarm levels in noisy work environments. The experiment is still ongoing, and the results of the study, including model performance, will be presented at the conference.

4. REFERENCES

[1] International Organization for Standardization (ISO), *ISO 7731 — Ergonomics — Danger signals for public and work areas — Auditory danger signals*, 2003.

[2] J. Żera and A. Nagórski, "Preferred Levels of Auditory Danger Signals", *International Journal of Occupational Safety and Ergonomics*, vol. 6, no. sup1, pp. 111-117, 2000.

[3] C. Laroche, C. Giguère, V. Vaillancourt, K. Roy, L.-P. Pageot, H. Nélisse, N. Ellaham and F. Nassrallah, "Detection and reaction thresholds for reverse alarms in noise with and without passive hearing protection", *International Journal of Audiology*, vol. 57, no. sup1, pp. S51-S60, 2018.

[4] L. Schell-Majoor, J. Rennies, S. D. Ewert, and B. Kollmeier, "Application of psychophysical models for audibility prediction of technical signals in real-world background noise", *Applied Acoustics*, vol. 88, pp. 44-51, 2015.

[5] T. Dau, D. Püschel, and A. Kohlrausch, "A quantitative model of ''effective'' signal processing in the auditory system. II. Simulations and measurements", *The Journal of the Acoustical Society of America*, vol. 99, no. 6, pp. 3623-3631, 1996.

[6] B. R. Glasberg and B. C. J. Moore, "Development and Evaluation of a Model for Predicting the Audibility of Time-Varying Sounds in the Presence of Background Sound", *Journal of the Audio Engineering Society*, vol. 53, no. 10, pp. 906-918, 2005.

[7] M. L. Jepsen, S. D. Ewert, and T. Dau, "A computational model of human auditory signal processing and perception", *The Journal of the Acoustical Society of America*, vol. 124, no. 1, pp. 422-438, 2008.

[8] J. Abeßer, "A Review of Deep Learning Based Methods for Acoustic Scene Classification", *Applied Sciences*, vol. 10, no. 6, 2020.

[9] Roy. D. Patterson, T. F. Mayfield, D. E. Broadbent, A. D. Baddeley, and J. Reason, "Auditory warning sounds in the work environment", *Philosophical Transactions of the Royal Society of London. B*, Biological Sciences, vol. 327, no. 1241, pp. 485-492, 1997.

[10] Roy. D. Patterson and R. Milroy, *Existing and recommended levels for the auditory warnings on civil aircraft*, MRC Applied Psychology Unit, Cambridge, Contract Report 7D/S/0142, 1979.

[11] Y. Zheng, C. Giguère, C. Laroche, C. Sabourin, A. Gagné, and M. Elyea, "A Psychoacoustical Model for Specifying the Level and Spectrum of Acoustic Warning Signals in the Workplace", *Journal of Occupational and Environmental Hygiene*, vol. 4, no. 2, pp. 87-98, 2007.

[12] The Freesound project, https://www.freesound.org.

[13] BigSoundBank by Joseph SARDIN, https://www.bigsoundbank.com.

[14] N. Turpault and R. Serizel, "Training sound event detection in a heterogenous dataset", in *Proc. DCASE Workshop*, 2020

[15] E. Cakir, S. Adavanne, G. Parascandolo, K. Drossos, and T. Virtanen, "Convolutional recurrent neural networks for bird audio detection," in *Proc. EUSIPCO*, 2017.

[16] F. Effa, R. Serizel, J.-P. Arz and N. Grimault, "Convolutional Neural Network for Audibility Assessment of Acoustic Alarms", in *Proc. DCASE Workshop*, 2022.