Applied Acoustics 88 (2015) 44–51

Contents lists available at ScienceDirect

Applied Acoustics

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

Application of psychophysical models for audibility prediction of technical signals in real-world background noise

Lena Schell-Majoor [a,*], Jan Rennies [a], Stephan D. Ewert [b], Birger Kollmeier [b]

[a] Fraunhofer IDMT, Project Group Hearing, Speech and Audio Technology, Marie-Curie-Str. 2, 26129 Oldenburg, Germany
[b] Medizinische Physik and Cluster of Excellence Hearing4All, Universität Oldenburg, 26111 Oldenburg, Germany

ARTICLE INFO

Article history:
Received 28 January 2014
Received in revised form 29 July 2014
Accepted 6 August 2014
Available online 30 August 2014

Keywords:
Psychoacoustics
Auditory models
Detection thresholds
Masking effects
Real-world signals
Audibility

ABSTRACT

A valid, objective computation of whether a real-world sound is detectable in a real-world acoustical environment is highly desirable in many noise control applications. However, most current prediction approaches have not been validated for this purpose and have not been tailored towards predicting the influence of certain signal features, such as the temporal structure or the spectral content of the masker or target. In order to evaluate the applicability of prediction approaches with respect to these signal features, detection thresholds of various real-world signals were measured for normal-hearing listeners. The detection thresholds depended on the temporal structure and spectrum of the target and the spectrum of the masker. The data were compared to predictions of five approaches ranging from time-averaged technical measures to psychoacoustic models, which incorporate these signal features to different extents. In general, the correspondence between predictions and the experimental data was better for the psychoacoustic models than for the results of the technical measures. Even though all models could account for most of the key effects in the experimental data, only the psychoacoustic models were able to predict the influence of the temporal structure of the signals. One of the models showed clear advantages in prediction performance, reaching an overall determination coefficient of $R^2 = 0.94$. This underlines the applicability of psychoacoustic models for correctly predicting audibility in real-world applications.

1. Introduction

The detection threshold of a sound in a given background noise is the level at which it is just detectable. Determining such detection thresholds of targets in the presence of background noise offers high potential for practical applications and sound quality engineering: On the one hand, the audibility of warning or indication sounds has to be ensured (see, e.g., [1,2]) and on the other hand, detection thresholds can serve as design goals for annoying or disturbing sounds which are desired to be imperceptible, such as the noise of wind turbines [3]. To this end detection thresholds can be measured in listening tests. In this case, each combination of a target and background noise requires an individual measurement, which is time consuming and costly. A cost- and time-efficient alternative is to predict detection thresholds based on perceptually relevant features of the target and the masker using psychophysically motivated auditory models of masking or loudness perception (e.g. [4,5]). While basic mechanisms underlying acoustic signal detection have been investigated in detail using artificial sounds such as pure tones or narrowband noises, detection thresholds of real signals have received considerably less attention. However, some approaches have been made in order to predict detection thresholds of real signals. A rather sophisticated model was developed by Glasberg and Moore [5] and evaluated for predicting the detection thresholds of mobile telephone ring tones in different background sounds. Similar model approaches were applied in other studies, e.g., to predict the masking effect of natural sounds on the noise of wind turbines [3] and to predict the detectability of exterior vehicle noise [6]. Additional psychoacoustic models exist, which attempt to reproduce the human hearing process and are derived from findings in psychoacoustic experiments [4,7,8]. They have been evaluated by predicting detection thresholds of artificial stimuli, such as tones in noise or modulation detection, and showed good performance [7,9].

The goal of the present study was to test the applicability of five prediction approaches for a broad range of real technical targets. Detection thresholds of various real-world signals were measured for normal-hearing listeners. In addition, modified steady signals with preserved long-term spectrum were used. Two prediction

---

* Corresponding author. Tel.: +49 441 2172 400.
E-mail addresses: Lena.Schell-Majoor@idmt.fraunhofer.de (L. Schell-Majoor), Jan.Rennies@idmt.fraunhofer.de (J. Rennies), Stephan.Ewert@uni-oldenburg.de (S.D. Ewert), Birger.Kollmeier@uni-oldenburg.de (B. Kollmeier).

http://dx.doi.org/10.1016/j.apacoust.2014.08.001
0003-682X/

Table 1
Overview of targets.

Targets	Origin
Car engine start[a]	Freesound.org
Car rattling	Supplement to [11]
Helicopter	Freesound.org
Bike bell	Freesound.org
Car window[a]	Freesound.org
Passenger aircraft fly-by	Freesound.org
Audible pedestrian traffic signal (APTS), standby[a]	Navtec.de
APTS active[a]	Navtec.de

[a] Indicate that stationary signals were generated based on those targets and used as additional targets.

approaches based on simple technical measures and three more complex approaches based on psychoacoustic models were used. Apart from their complexity, the models also differed with respect to the signal features used to predict detection thresholds. The technical approaches were based on simple technical features of the signals such as the level or the spectral content. The psychoacoustic models analyzed level, spectral content and the temporal information (modulation features) of the signals. Hence, they used an additional signal feature, the temporal structure, and also analyzed the combination of the different signal features with regards to processes and properties of human auditory perception (e.g., frequency selectivity, spectral and temporal masking mechanisms). Therefore, one could expect the psychoacoustic models to predict detection thresholds of targets with higher accuracy than the simple technical approaches. The goal of comparing prediction approaches was to investigate which signal features are crucial and how much computational effort is necessary to predict detection thresholds with reasonable accuracy. Model performance was assessed by comparing the prediction results to experimentally measured detection thresholds. It was investigated which features of the target and background signal have to be analyzed in order to get the closest match to the data. As the technical approaches were based on long-term power-spectra, while the psychoacoustic models also considered temporal features, the role of modulation features for real-world signals was assessed. The general goal of this study was thus to compare the effectiveness of technical and perceptual models for predicting thresholds of real-world signals.

2. Experiments

2.1. Method

2.1.1. Subjects[1]

Nine subjects (5 male and 4 female subjects) aged between and 23 and 31 years with previous experience in listening tests participated voluntarily in this study. They were not paid for their participation. All subjects reported normal hearing abilities and had pure-tone thresholds of less than or equal to 20 dB HL at audiometric frequencies in the range 125 Hz–8 kHz.

2.1.2. Stimuli

Recordings of eight targets, listed in Table 1, and two background sounds (maskers) were used. Stimuli were taken from data bases (as indicated in the second column of Table 1) and had a sample rate of 44.1 kHz except for two targets (Audible Pedestrian Traffic Signal; APTS and APTS-on used in German pedestrian traffic lights to acoustically signal a “stop” or “go” for visually impaired pedestrians). These were up-sampled from an original sample rate of 8 kHz. The first masker was recorded inside a small car going at approximately 140 km/h (own recording, see [10] for details). The second masker was rain on a roof (taken from Freesound.org). While both maskers had a rather stationary temporal envelope, they differed considerably in their power spectra: the car noise had most of its energy in the low-frequency range while the spectrum of the rain noise included more high-frequency components. This difference was used to evaluate the role of the spectral content of the background noise for measured and predicted detection thresholds. Similarly, the spectra of the targets varied markedly: some of the targets had a broadband spectrum (e.g., the starting car engine or the aircraft), while other targets had narrowband spectra centered in different frequency regions. The helicopter sound, for example, was a low-frequency sound, whereas the APTS-on was mainly characterized by distinct high-frequency components. Third-octave (audio) power spectra and third-octave envelope power spectra for all stimuli are shown in Fig. 1. The (Hilbert) envelopes were calculated by taking the absolute values of the analytical signal. As only fluctuations in the envelope are of interest, the mean value of the envelope (DC) was removed. Afterwards, third-octave power spectra were calculated. The targets also varied markedly in their temporal structure. While the aircraft signal showed a rather stationary temporal envelope, targets like the engine start and bike bell were characterized by strong amplitude modulations. This large variety of maskers and targets was intentionally included in the stimulus set in order to cover a broad range of potential application scenarios and to test if prediction models could deal with the large variety of physical properties. To further investigate the influence of the temporal structure of the targets on detection thresholds, four additional targets were generated. The temporal envelope fluctuations were eliminated for a subset of four of the eight original targets (marked in Table 1) by generating stationary signals with identical long-term power spectra. This was accomplished by shaping the spectrum of white noise with a customized filter according to the long-term spectrum of the original target and adjusting the root-mean-square (RMS) level. All targets had a length of 1 s and were excerpts from longer signals. The targets were temporally centered in the maskers that had a length of 1.5 s and were presented at a fixed level of 75 dB SPL.

2.1.3. Procedure and apparatus

Detection thresholds were measured using monaural presentation in three sessions using a three-interval, three-alternative forced-choice (3-AFC) paradigm in the AFC framework for MATLAB described in [12]. Sounds were presented via headphones (Sennheiser HD 650) in a sound-attenuating booth. Each interval contained the masker but only one randomly selected interval included the target. The masker was identical within a set of three intervals (trial). For each trial, the masker was randomly cut from a long-duration signal, resulting in independent maskers in the different trials. The task was to choose the interval containing the target. If the subjects did not perceive the target in any of the intervals they were instructed to guess. The intervals (1.5 s duration) were separated by 0.5 s of silence. The level of the target was varied depending on the subjects’ responses following a one-up, two-down procedure starting with a level of 50 dB SPL. The initial step size was 8 dB. The step size was halved after every second reversal, until a step size of 1 dB was reached. At this step size (measurement phase), the measurement was continued for another 8 reversals. The average signal level at these eight reversals was taken as the detection threshold. The one-up, two-down procedure converges on the 70.7%-correct level of the target [13]. One measurement was performed for each subject and signal. The presentation order

---

[1] This study has been carried out in accordance with The Code of Ethics of the World Medical Association (Declaration of Helsinki) for experiments involving humans.