# RECOGNIZABILITY AND PERCEIVED URGENCY OF BICYCLE BELLS

*Lisa Frohmann, Marian Weger, Robert Höldrich*

Institute for Electronic Music and Acoustics (IEM)
University of Music and Performing Arts
Graz, Austria
li.frohmann@gmail.com

## ABSTRACT

Raising awareness about how alarm sounds are perceived and evaluated by an individual in traffic scenery is important for developing new alarm designs, as well as for improving existing ones. Bearing a positive contribution to road safety, cyclists and pedestrians especially can benefit from appropriate alarming bell and horn sounds. Primarily, the alarm signal should evoke a precise idea of what is the source of the warning and the desired reaction to it. Furthermore, it should not be masked by other noises thus going undetected by the ear. Finally, an appropriate warning signal should transmit the urgency of a given situation, while at the same time, it should not cause other road users and pedestrians to startle.

In two listening experiments, we examined the perception of commonly available bicycle bells and horns. Average typicality or recognizability as a bicycle bell among other everyday sounds has been investigated through a free identification task. In a second experiment, we tested perceived urgency of the warning sounds in relation to traffic noise. This article further provides a survey on non-verbal alarm design, as well as an analysis of acoustic properties of common bicycle bells and horns. Consequently, a linear regression model presents the relationship between named properties and perceived urgency.

It is our intention to give an insight into the often unattended but important issue of the perception of auditory warning sounds in our everyday acoustic environment.

## 1. INTRODUCTION

Acoustic warning signals are part of our everyday lives, designed to make hazardous environments and situations safer by giving accurate information about an expected risk at the appropriate time. We are confronted with such alarm signals in various situations - from road traffic, aviation monitoring and countless industries, to medical equipment in the hospital. Whether we hear the honk of a car horn or the constant beeping of an electrocardiogram monitor (ECG), all these sounds have a commonality - they bear an important message for their target and they manage to draw attention to an incident in a preferably short period of time.

In contrast to warnings using speech, non-verbal auditory alarms lack the feature of precise expression of what they are informing about and what the desired reaction to this alarm should look like. Therefore, it is important to take into account temporal as well as spectral signal properties in order to design the best possible warning sound for a specific scenario. Originally established for alarm systems on military aircraft [1], several acoustic properties form criteria for adequate acoustic warning that suit all application areas. As proposed, warning sounds...

- must be unique in the noise environment,
- must be discriminable from other sounds of the surrounding,
- should convey the correct relative urgency for the associated priority level,
- should be presented at correct audio level for reliable detection.

Edworthy et al. [2] emphasize on the term of *Urgency* which makes a ranking of warning sounds possible. Its understanding is crucial for the success of alarm design:

> *"[...] there is often a serious mismatch between the perceived (psychoacoustic) urgency of a warning -its implicit urgency as a function of its sound parameters- and its situational urgency -the degree of urgency that the operator (e.g. pilot, nurse) has learned to associate with the warning as a function of the situation itself."* [2]

The authors studied the effects of different signal parameters on perceived urgency of warning sounds. These parameters were based on an approach by Patterson [3] who constructed sound-bursts from impulse-trains which were modified in pitch, tempo, intensity, melody, and rhythm. Participants of a listening experiment then rated and ranked the perceived urgency of these synthetic alarms under laboratory conditions, leading to the following conclusions: The higher the pitch, the more irregular the harmonics, and the faster the pulse rate, the greater the perceived urgency [2].

These findings, however, could not be approved clearly for real, already existing alarms and for subjects being under high workload when alarmed. On the one hand, there seems to be a connection of acoustic characteristics and urgency on a low level. On the other hand, there has to be a higher level influence of auditory learning and acculturation. The perception of urgency (and therefore potential danger) is linked to a mental representation of the cause [4]. Without existing knowledge or education around the meaning of any specific alarm, it is difficult to find an appropriate reaction to it and the alarm may have a startling effect and lead to confusion.

A lot of research concerning intelligent alarming has been done in the fields of medical care, as well as aviation and automotive industry, e.g., [5, 6, 7, 8]. New car horn designs and electric vehicle noises that match the road users' needs and the designers' intentions concerning branding in equal shares, have been explored [9]. As modern cars are built to be as soundproof as possible, new problems arise, and solutions have to be found to make ambulance alarms and police sirens audible again. Inventions such as additional low-frequency sirens [10], that penetrate the hard surfaces of the car body, are dealing with the situation. Bringing the warning directly inside the car via Radio Data System (RDS) yields another approach [11, 12].

Other road users such as cyclists and pedestrians are unjustly excluded in these developments. Alarm instruments that warn cyclists, as well as sounds that cyclists use for signaling others, are an important safety feature compulsory by law in most countries. However, these mandatory safety tools are poorly developed, unused and misinterpreted on a daily basis. The sound of a bicycle bell is often associated with a low level of threat or even with pleasure, which certainly interferes with the effectiveness of the warning [13].

In this article, we present two listening experiments in which commonly available bicycle bells and horns have been examined on the basis of the acquired knowledge about properties and design principles of auditory warning sounds. As the mental representation of an alarm's cause is of great importance to the reaction to it, a classification test was performed in order to reveal if the tested bicycle warnings could be identified as such among a set of everyday sounds (Sec. 2). Furthermore, the level of urgency that conventional bicycle bells carry to their target is of interest, as well as if an urgency ranking of different bell types can be obtained. Therefore, a second listening experiment on perceived urgency of bicycle warning sounds has been carried out (Sec. 3). Spectral and temporal properties of the bell sounds (perceived pitch, spectral centroid, roughness, onset time and relative intensity) are examined in Sec. 4 and set in relation to the measured urgency values, yielding an objective model formed by linear regression (Sec. 5). A general discussion and conclusion finally follows in Sec. 6.

## 2. EXPERIMENT 1: RECOGNIZABILITY TESTING

The aim of this listening experiment was to give information about how well the traditional bike bells as well as state-of-the-art bicycle warning tools were identified and accepted as such.

### 2.1. Procedure and Apparatus

*Subjects.* 17 subjects participated in Experiment 1: 5 females and 12 males with an average age of 24.5 years (ranging from 21 to 29). All of them were students in Audio Engineering at Graz University of Technology and Graz University of Music and Performing Arts. Nobody reported a hearing difficulty, and 15 of 17 had experience with listening experiments.
*Apparatus.* The experiment took place in an acoustically treated room of 4.3 m × 6.2 m × 3.4 m size (w × l × h), with reverberation times between 0.15 and 0.22 s in the relevant frequency range above 200 Hz. The technical setup included a Genelec 8020 CPM studio monitor connected to a Samsung 700Z laptop through an RME Fireface UCX audio interface. The stimuli were played back under Windows with Cubase 8 software. Subjects were seated in the center of the room with the loudspeaker located in a distance of 1.76 m in line of sight.
*Procedure.* The subjects were asked to freely identify 20 everyday sounds which they could encounter on the streets of a city. Before starting the experiment, subjects were informed vaguely about this imagined scenario without revealing the exact stimuli. Each stimulus was presented twice, with a short pause in between. Then, before continuing with the next stimulus, there was a pause of about 5 seconds for writing down the answer. Subjects were not allowed to pause or stop the experiment, as their first impression and classification was desired. The order of the stimuli was randomized between subjects in order to avoid a possible influence of order.
*Stimuli.* The 20 stimuli were selected to include only 10 actual bicycle warning sounds. These common bell stimuli included two large (l.) and two medium sized bells that have to be flicked horizontally, i.e., twisted (TB), three bells with a hammer stroke mechanism (H), a bike horn, and the electric bicycle bell *Hornit* by Bullet Ventures Limited [14], which offers two sounds: a loud piercing siren for street use and a horn-like sound for more quiet settings such as cycling through a park. The other half of the stimuli consisted of 8 acoustic warning signals (2 car horns, 2 acoustic traffic lights for the blind, the ringing of a mobile phone, a tramway bell, a doorbell, and a whistle), and 2 miscellaneous sounds (the rattling sound of a keyring and a wind chimes sound). The latter sounds were included in the hope of making the experiment more interesting and less obvious to the participants. The stimuli had a duration of 3 to 6 seconds.

### 2.2. Results

Table 1 shows the answer matrix for the classification test in a reduced version. Sounds that were correctly identified by most of the subjects (e.g., tramway signal, rattling of a key ring, etc.) were excluded from the further evaluation as they were not the target of our examination but were included only for experiment design reasons.

The individual answers were grouped under umbrella terms. From these results, the recognizability as a bike bell was calculated for each stimulus as the percentage of answers which included both 'bicycle' *and* 'bell/horn'. The resulting values of bicycle typicality are shown in Tab. 1. 95 % confidence intervals are added, assuming a binomial distribution.

A hierarchical clustering analysis of the classification results revealed three main clusters: Cluster A consists of both large 'TB' type bells. Cluster B contains the small 'TB' bells as well as the 'H' type bells. Cluster C combines the bicycle horn and the two Hornit sounds. For further analysis we additionally split Clusters B and C in order to achieve five clusters which can even be formed instinctively by the physical mechanism of sound generation and thus typical auditory gestalt: (1) 'Ding-dong', (2) 'Ding', (3) 'Rrring', (4) 'Honk' and (5) 'Electronic'.

### 2.3. Discussion

The results of Experiment 1, investigating recognition as bicycle and clusterization, matched our assumptions made beforehand. Well known, most common and purchasable bell types such as hammer bells and small to medium sized twist bells with a rattling sound reached high bicycle typicality values. Larger twist bell models gained average to low scores, and the new Hornit bell tool did not create a mental connection of a bicycle for neither of the subjects. The low values of bells of group (1) (cluster B) could be explained by their resemblance to a doorbell and therefore by assignment to a different setting and cause. Tab. 1 displays this

| | | Classification category | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| group | Stimulus | bicycle bell/horn | car horn | mobile phone | general horn | doorbell | bird chirp | general bell (small) | general bell (large) | toy sound | alarm/siren | other | Bicycle typicality in % | CI (95 %) |
| 1 | TB l. yellow | 5 | | | | 10 | | | 2 | | | | 29 | 10-55 |
| | TB l. pink | 8.5 | | | | 6.5 | | 1 | 1 | | | | 50 | 27-77 |
| 2 | H lilac | 13 | | | | | | 3 | | | | 1 | 76 | 50-93 |
| | H black 1 | 12 | | | | | | 1 | 1 | | | 3 | 71 | 44-89 |
| | H black 2 | 7 | | 1 | | | | 3 | 2 | | | 4 | 41 | 18-67 |
| 3 | TB Dinotti | 16 | | | | | | 1 | | | | | 94 | 71-99 |
| | TB Puch | 16 | | | | | | 1 | | | | | 94 | 71-99 |
| 4 | Bicycle horn | 7 | | | 9 | | | | | 1 | | | 41 | 18-67 |
| 5 | Hornit Park | | 2 | | 7 | | | | | 3 | 3 | 2 | 0 | 0-19 |
| | Hornit Street | | | | 1 | | 5 | | | 7 | 2 | 2 | 0 | 0-19 |
| - | Doorbell | | | | | 10 | | 1 | 5 | | 1 | | 0 | 0-19 |

Table 1: Classification of stimuli in reduced form and calculated bicycle typicality with confidence interval (CI).

mismatch. Yet interestingly the original doorbell sound was never inversely mistaken for a bike bell. The bike horn with a recognizability of 41 % was often classified either as a typical children's bike horn or a common horn with no relation to a bicycle at all. This supports our personal impression that such bike horns are rare on adult bicycles, at least in the cultural environment of the tested participants.

The grouping of bell types will be used in Experiment 2 for the purpose of comparing types rather than individual bells. Considering the fact that the small groups (1) to (5) can be obtained by further division of clusters A and C, we will stick to the five small groups, even if they are not completely arguable perception-wise through cluster analysis. Most importantly, however, they give an adequate representation of bell types available on the market.

## 3. EXPERIMENT 2: URGENCY

### 3.1. Procedure and Apparatus

*Subjects.* The 17 subjects to participate in Experiment 2 were the same as in Experiment 1.
*Apparatus.* The experiment was conducted in the same room as the first experiment. The technical setup comprised four Genelec 8020 CPM studio monitors in the corners of a 3.53 m × 3.53 m square, connected to a Samsung 700Z laptop through an RME Fireface UCX audio interface. The experimentation software was implemented in Matlab, with audio playback through Pure Data (Pd). Subjects were seated in the center of the square, with the speakers in a distance of 2.45 m to the center of the head, in 1.2 m height, in azimuth angles of 45°, 135°, 225°, and 315°. Calibration was done with an NTI Audio M2210 measurement microphone and Brüel&Kjaer acoustical calibrator Type 4231.
*Stimuli.* Subjects were presented 12 bicycle bell and horn sounds. 10 of these were the same as in Experiment 1. The two additional ones were slightly modified versions of Stimuli 'TB large pink' and 'H black 2'. In particular, the spectrum of 'TB large pink' was altered by boosting the gain at 2.5 kHz, 7 kHz and 11.8 kHz by 12 dB in order to create a more piercing sound. Stimulus 'H black 2' was reduced in tempo, equivalent to a stimulus duration stretch from 3 to 6 seconds.
For the condition with background soundscape, different field-recordings of environmental noise at typical street crossings in the city of Graz were layered to form an indefinite and blurry version of a typical urban acoustic environment. The traffic noise was played back quadrophonic with all four loudspeakers; the sound pressure level measured at the listening position was set to match the one at an average street crossing.
In order to ensure an ecologically valid presentation of the stimuli, their sound pressure levels were adjusted to match the original ones as close as possible. In a first step, the sound pressure levels of two prototype bell stimuli – a traditional one and the horn – were calibrated to match the real ones at the loudspeaker position by comparing the levels measured with the measurement microphone at approximate head position. All other stimuli were then subjectively adjusted to match the amplitudes of these prototype bells.
*Procedure.* Experiment 2 was divided into two multi-stimuli tests. For part one, subjects had to blindly rate the urgency of all the stimuli over a Matlab-based GUI from 'not urgent' to 'very urgent' using horizontal sliders without scaling. The direction of sound playback differed pseudo-randomly over the four loudspeakers. Subjects were able to listen to all stimuli in their preferred order, enabling them to compare stimuli pairwise and also sort their answers. In part two, the urgency rating experiment of part one was repeated with the additional traffic soundscape, simulating real life conditions. As experiments excluding the acoustic environment of the warning signal have already been carried out [2], we were interested in testing both the noisy and the silent scenario for the bike bells to detect differences in urgency ratings caused by the context of traffic noise.

### 3.2. Results

Figures 1 and 2 show boxplot diagrams for the urgency ratings of the 12 individual bell stimuli in the two different conditions without and with background noise, with the mean values (bold black line) displayed additionally to the median (narrow line).
We proceed with a statistical analysis of these results. In both conditions, some outliers were identified by Grubbs' test and thus replaced by the mean value of the affected stimulus (3 stimuli of 1 participant in part 1, 2 stimuli of 2 participants in part 2). Lilliefors test could not reject the null hypothesis of normally distributed data at a significance level of $p<0.05$ for the majority of the data (25 % non-normal for stimuli without background noise, 8 % non-normal for stimuli with background noise). We therefore performed a two-way (Stimulus × Noise) analysis of variance (ANOVA).

ANOVA showed a highly significant main effect of stimulus type on urgency rating at the $p<0.05$ level ($F(11,176)=15.69$, $p<0.001$). This indicates that different bell types incorporate significantly different perceived urgency. There was also a significant main effect of traffic noise on urgency rating ($F(1,16)=4.50$, $p=0.034$). Furthermore, the interaction between stimulus and traffic noise was highly significant ($F(11,176)=3.13$, $p<0.001$).

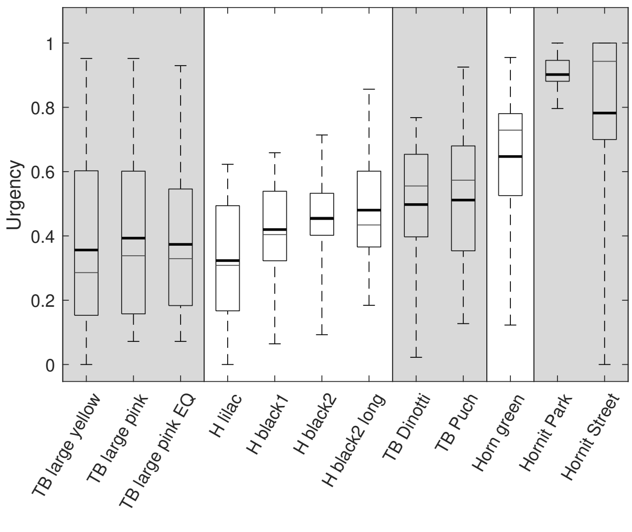



Figure 1: Urgency rating without background traffic noise. Alternating background filling indicates membership to the bell groups of Fig. 3.

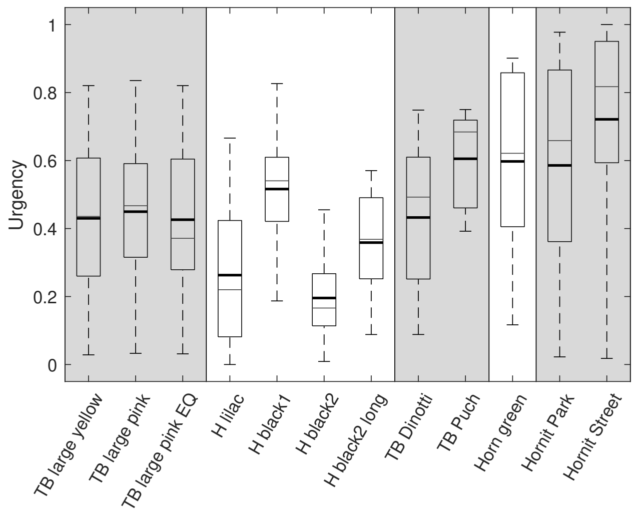



Figure 2: Urgency rating with background traffic noise. Alternating background filling indicates membership to the bell groups of Fig. 3.

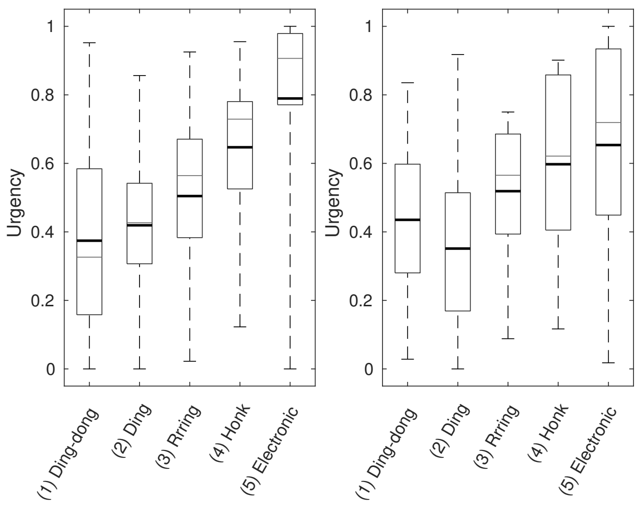
In order to compare the individual stimuli's urgency values to each other, pairwise t-tests were performed. The differences between stimuli scores were examined averaged over background noise conditions. The Hornit type bells were not significantly different to each other, while they were significantly more urgent than all other bells ($p \leq 0.007$), except for the bicycle horn which did not differ significantly from Hornit Park ($t(16)=-1.808$, $p=0.089$), but showed significantly lower urgency compared to the siren-like Hornit Street ($t(16)=-2.454$, $p=0.026$).

The bicycle horn was rated significantly more urgent than the large TB bells ($p \leq 0.002$), as well as significantly more urgent than the hammer type bells ($p \leq 0.035$). However, the difference in urgency to the TB type bells of rattling sound character was not significant. Furthermore, the three large TB type stimuli were not significantly different to each other. All of these, however, were significantly less urgent than the clattering 'TB Puch' and horn stimuli ($p \leq 0.005$). Pairwise comparison between the two medium sized rattling twist bells revealed a significant disparity in urgency, meaning 'TB Puch' was significantly more urgent than 'TB Dinotti' ($t(16)=2.312$, $p=0.034$).
Inside the group of 'H' type bells, 'H black1' and 'H black2 long' were both significantly more urgent than the other two stimuli ($p \leq 0.037$), while other pairwise comparisons were not significant. Both 'H lilac' and 'H black2' were also significantly less urgent than the two rattling 'TB' bells ($p \leq 0.009$). In addition, 'TB large pink' was significantly more urgent than 'H lilac' ($t(16)=2.317$, $p=0.034$), and 'TB Puch' was significantly more urgent than 'H black1' and 'H black2 long' ($\leq 0.027$). All combinations that have not been named so far, were not significant concerning perceived urgency.

With concern to the significant effect of background noise on perceived urgency, a pairwise t-test showed that average urgency scores were significantly lower in a noisy scenario (mean=0.47) than without noise (mean=0.51) ($t(16)=2.343$, $p=0.032$).

In a next step, we also performed a statistical analysis of the pooled results in the five groups as derived in Experiment 1. Boxplots for both background noise environments are shown in Fig. 3.

No outliers were detected by Grubbs' test for neither of the traffic noise conditions. Lilliefors test, however, yields 60 % non-normal data in quiet surrounding, but 0 % non-normal data in the noisy setup. We will therefore rely on the two-sided Wilcoxon rank-sum test in the following statistical analysis.




(a) without traffic noise (b) with traffic noise

Figure 3: Urgency ratings of the bell type groups, without (a) and with (b) background traffic noise.

The noise condition had a significant effect only for stimulus group (5): The Hornit type stimuli were significantly more urgent without (mean=0.789) than with background noise (mean=0.654) ($p=0.036$). Statistical analysis of pairwise rank-sum tests, averaged over noise condition was performed between the respective stimuli groups and can be summarized as follows.

Groups (4) and (5) did not show a significant difference in urgency rating (p=0.103), but both were perceived significantly more urgent than all other groups ( (3) vs. (4): p=0.026, all other combinations: p≤0.001). Similarly, groups (1) and (2) were not significantly different concerning perceived urgency (p=9.912); however, both were perceived significantly less urgent than all other groups ( (1) vs. (3): p=0.007, all other combinations: p≤0.001).

The interaction between stimulus group and noise condition could be explained through the following observation: According to Wilcoxon rank-sum test, in the noisy condition, group (1) was significantly more urgent than group (2) (p=0.039, with means 0.435 and 0.351, respectively). In contrast to this, it was yet the other way around without noise (means 0.374 and 0.419). However, the latter observation was not significant (p=0.191). All other pairwise comparisons maintained the sign of their urgency difference throughout both noise conditions.

### 3.3. Discussion

Interpreting the obtained results, we can state that traffic noise is responsible for a drop of the average rating of urgency. One explanation could be that the additional traffic noise may generate a level of stress for the subjects and takes away the focus from the bell stimuli to other potentially dangerous sound sources such as cars – the bicycle bell sounds are perhaps put in relation to the possibly more urgent traffic noise. Somehow connected to this assumption, possibly more important information gained from traffic noise, might induce informational masking effects on the examined bell sounds. Another reason could be masking of spectral components of the bell sounds by the traffic noise, thus reducing their perceived urgency.

However, the noise condition affected urgency significantly only for the Hornit sounds which were the overall winners of the urgency rating experiment. Interestingly, without being asked, the majority of the subjects reported confusion, when having to rate the Hornit sounds (especially the Park mode, designed by the manufacturer for calm environments, hence aiming at lower urgency). Despite giving high values in the no-noise test, participants did not sense an urge to react for the same stimulus in noisy environment, stating that it sounded too far away to be of concern.

The 'Rrring' type bells of group (3) seem to form a stable class of average urgency, as the difference between noise conditions is rather small. We can see a similar picture for the bike horn, but with a greater spreading of subject ratings when having to judge urgency in background noise. That might be because the horn is often associated with a toy sound or a children's bike bell, as well as because in the noisy setting, direction and source of the sound are not easily assigned.

The hammer bells of group (2) got lowest ratings in noisy environment, which seems logical as these consist of one pulse only and appear weak and insufficient for the environment. The traffic noise seems to absorb important features of the sound, which are available to the ear in silent surroundings, with the effect of lowering perceived urgency when these acoustic cues are less prominent due to spectral masking.

## 4. ACOUSTICAL DESCRIPTORS OF URGENCY

A spectral and temporal analysis of the given bell sounds in relation to the urgency estimation aimed to explore possible correlations. Given that the bell types incorporated a high grade of variation in sonic characteristics, adequate signal analysis methods are desired in order to predict the urgency of the recorded bicycle bells from acoustical descriptors. We decided for five common signal attributes: perceived pitch, spectral centroid, mean relative roughness, onset time, and relative signal power. These attributes can be easily calculated or at least roughly estimated for any recorded bell sound with existing models and methods, illustrating both temporal (e.g., onset) and spectral (e.g., pitch, centroid) components which relate to perceived urgency.

### 4.1. Perceived Pitch

Estimating a precise value for perceived pitch of a bell sound or a burst of sound containing numerous frequencies is problematic. For many musical instruments such as cordophones or aerophones, it is easy to identify a fundamental frequency, but for said bells, chimes or percussive sounds, i.e., idiophones, the situation of finding pitch is remarkably unclear [15]. Evaluation of the spectrum does not always lead to the right solution, as the perceived pitch can be influenced by difference tones and virtual pitch suggested by the human brain. Despite these problems, there is no doubt that one instantly gets some impression of pitch, when hearing a bell sound, even if this pitch differs between individuals. A higher fundamental frequency does affect the sensing of urgency, although studies like [2] (and others cited therein) describing it as a minor influence.

To explore the case for bicycle bells, perceived pitch was measured by means of a frequency-matching experiment performed by the investigator, i.e. through adjustment of a pure sine wave to match the perceived pitch of the individual bell stimulus. For sounds that consist of two or more consecutive or simultaneous auditory events, the value refers either to the first event or to the overall impression on the sound respectively. As this is a very subjective approach, readers are warned that perceived pitch might differ for a larger pool of subjects conducting such a frequency-matching experiment. Perceived pitch is displayed in Tab. 2.

### 4.2. Spectral centroid

The spectral centroid is an indicator of the spectrum's "center of mass", its location in frequency influencing the perception of a sound's brightness. Incorporating a significant amount of signal energy at high frequencies, we assume that a bright sound with a centroid in the human ear's most sensible frequency range from about 2 to 5 kHz could favor alertness and induce higher perceived urgency.

The spectral centroid is generally calculated – following Eq. 1 – as the weighted mean of the frequency bins of a signal obtained by Fourier Transform (with $\frac{kf_s}{N}$ being the frequency of bin $k$ ), weighted with the respective magnitudes $|X(k)|$ for each bin. For this experiment's computation, the bell sounds were additionally weighted with an A-weighting-filter before spectral density estimation was performed using Welch's method (`pwelch` in Matlab). Spectral centroid values are displayed in Tab. 2.

$$Centroid = \frac{\sum_{k=0}^{\frac{N}{2}-1} \frac{kf_s}{N} |X(k)|}{\sum_{k=0}^{\frac{N}{2}-1} |X(k)|} \tag{1}$$

| | urgency rating | | signal parameter | | | | |
|---|---|---|---|---|---|---|---|
| Stimulus | with noise | without noise | perceived pitch kHz | spectral centroid kHz | mean roughness | onset time ms | signal power relative to loudest $\Delta$ dB(A) |
| Hornit Street | 0.72 | 0.78 | 1.28 | 2.95 | 6.53 | 82 | -1.4 |
| TB Puch | 0.61 | 0.51 | 0.91 | 6.41 | 0.89 | 21 | -7.5 |
| Bicycle horn green | 0.60 | 0.65 | 0.49 | 3.14 | 0.36 | 15 | -3.3 |
| Hornit Park | 0.59 | 0.90 | 0.56 | 2.70 | 1.12 | 71 | 0.0 |
| H black 1 | 0.52 | 0.42 | 1.67 | 4.23 | 0.01 | 2 | -10.3 |
| TB large pink | 0.45 | 0.39 | 0.60 | 4.75 | 0.07 | 5 | -6.7 |
| TB Dinotti | 0.43 | 0.50 | 1.04 | 8.06 | 0.27 | 24 | -13.1 |
| TB large yellow | 0.43 | 0.36 | 0.59 | 3.61 | 0.12 | 3 | -5.6 |
| TB large pink EQ | 0.43 | 0.37 | 0.59 | 3.96 | 0.20 | 6 | -6.2 |
| H black 2 long | 0.36 | 0.48 | 1.34 | 8.87 | 0.00 | 17 | -13.4 |
| H lilac | 0.26 | 0.32 | 1.58 | 3.38 | 0.01 | 2 | -13.8 |
| H black 2 | 0.20 | 0.45 | 1.34 | 10.19 | 0.01 | 3 | -13.6 |

Table 2: Signal parameters and urgency rating. Stimuli are sorted by urgency rating with background noise in descending order.

### 4.3. Roughness

The property of roughness, as an effect of interaction of sinusoids close in frequency, is selected due to its potential of causing the “unpleasantness” to the ear. It might correlate to findings in [2], that inharmonic partials result in higher urgency ratings. In addition, most people urge to stop an annoying sound when it becomes too irritating, which could lead to a shorter reaction time for a high-roughness warning.

The roughness or sensory dissonance of A-weighted bell sounds was computed in Matlab using the `mirroughness` function of the *MIRtoolbox* [16]. This particular algorithm was implemented based on proposals of Plomp and Levelt for an estimation of roughness related to the beating phenomenon of sinusoids close in frequency to each other and Sethares' model of sensory dissonance based on their work. Latter is obtained by computing the spectral peaks of a sound and taking into account all the dissonances between all possible pairs of peaks, as cited in [16] and the MIRtoolbox manuals. For further information about the model, please consult Appendix G of [15]. As the computed roughness of the stimuli is a function of time, the mean roughness of the bell sounds is calculated in a last step (see Tab. 2).

### 4.4. Onset time

Onset time could also be a relevant factor affecting urgency perception. Experiments carried out by [2] and [4] led to the conclusion that a slow onset is considered less urgent than the stated standard onset of 20 ms for synthesized alarm sequences. One explanation could be that a fast onset implies a quick approach of a moving sound source towards the test subject with a certain intention, contrary to a slow onset probably not bearing the same level of exigence. Furthermore, the output over the total duration is obviously higher for such short-onset stimuli [2].

In the case of our complex stimuli, onset time (or attack time) was defined as the time it takes for a sound's envelope to rise from zero to the first maximum on the time axis. The temporal profile of each bell sound was plotted and onset time was measured by hand for each stimulus individually by the investigator. For sounds consisting of two or more consecutive pulses, onset time was measured only for their first pulse accordingly. The measured onset times (in ms) are displayed in Tab. 2.

### 4.5. Signal power

As mentioned before, stimulus intensity was not the focus of our investigation and we therefore tried to neglect this factor, while it is undoubtedly one of the strongest known factors having an impact on urgency, as stated by [6] and others cited therein. Relative balancing of the sounds was done by ear by two investigators for said reason. Despite that, the actual signal power reaching the subjects ears at the center of the experimental chamber was examined. Regardless of all previous arrangements, signal power could have had an influence on the subjects' rating preferences. Calculations of signal power were performed on the A-weighted bell sounds. Signal power $\Delta$dB(A), relative to the loudest stimulus ‘Hornit Park’, is displayed in Tab. 2.

### 4.6. Discussion

Evaluating the calculations of the sound properties introduced in the previous sections aimed at finding simple parameters for any type of bicycle bell stimulus, that could render prediction of urgency ranking possible or at least give a hint on what could be a valid factor affecting urgency besides sound intensity.

For the bell sounds' perceived pitch, no reasonable connection to the urgency values could be found at first sight. Bells within one group – which incorporate similar mechanisms of sound production and bell sizes – possess similar perceived “fundamental frequencies”. In between the bells of a bell group, there seems to be a small positive effect, but there is no overall relationship for the individual stimuli. Experiments in [2] did find some minor relation of fundamental frequency and urgency, though it does not seem to play a big role in the perception and rating process, especially if other salient spectral and temporal attributes are determined. The presence of other strong signal characteristics might be the reason or at least an influencing factor to our results.

As stated in Sec. 4.2, a higher spectral centroid implies higher brightness of a sound and might support the perception of a stimulus' message in terms of urgency. However, for bells with a spectral centroid above 6 kHz, urgency seems to decrease in both noise conditions. Below this line though, there is a great spread in urgency ratings. The Hornit bell sounds and the bike horn, with spectral centroids around 3 kHz, achieved far higher urgency scores than the large twist bells and two of the hammer bells with centroids in the range of about 3.3 to 5 kHz. Our assumption lacks an explanation for said low values, hence it is not possible to draw a conclusion for urgency behavior from the spectral centroid analysis only.

A high value of roughness of a bell sound corresponds clearly to a high urgency rating for – in descending roughness order for traffic noise condition – the Hornit sounds, TB Puch, the bike horn and TB Dinotti (both twist bells are of group (3) with a rattling character to the sound). In this observation, hammer bells and large twist bells indicate very small roughness values compared to the other groups. Interestingly, without traffic noise, hammer bells appear slightly more urgent than the large twist bells, despite holding the lowest average roughness scores. Subjects confusing the large bike bell with a doorbell sound might be a simple explanation for this order of the stimulus groups. In a noisy environmental context, however, the two named groups switch places, leaving most hammer bells on the bottom of the roughness scale.

The results for onset time obtained from our experiments could not support the findings mentioned in Sec. 4.4. Slow onsets for both Hornit bell sounds did not correlate with low, but with extremely high urgency scores ($1^{st}$ and $4^{th}$ place). The rattling TB Puch with an onset close to the standard value of 20 ms even reached the second highest rating on the scale. The bike horn with a faster onset gained 0.6 in traffic soundscape, while all other bells with the fastest onsets around 5 ms are piled on the bottom of the scale. It seems that other properties than the temporal approach of onset time are responsible for the rank obtained by the individual stimuli's ratings.

Finally, signal power estimation was related to the urgency ratings. There is a tendency for bell sounds of high signal power to be considered more urgent than ones of low power values. The large bell sounds again are the exception to this finding in the setting without background noise. Despite having higher power estimates than the group of hammer bells, urgency ratings appear too low as to fit the assumptions. Again, the mismatch of mental representation of sound source could be a reason for this, similar to the observations concerning roughness. For urgency ratings with background noise, there is a better fit of the large TBs in the power-urgency relation. Despite some changes in stimuli ranking due to other properties interfering, signal power certainly has great influence on urgency rating.

## 5. OBJECTIVE URGENCY MODEL

Following the computation of the signal parameters, we are interested if it is possible to model the relationship between urgency and the explanatory variables using multiple linear regression. All computations were performed in Matlab using the Statistics toolbox. As it is reasonable for psychoacoustic application, logarithmic scaling is used for the regressors pitch, centroid and roughness. Urgency rating with background soundscape is selected as the dependent and ecologically valid variable, representing a real-life traffic situation. The corrected Akaike Information Criterion for small sample sizes is applied for deciding which regressors are used in the model to avoid over-fitting. As we suspected, onset time and spectral centroid are not among the chosen parameters. The case of regressors roughness, pitch and signal power (loudness) is displayed in Fig. 4. All variables show positive coefficient values, thus meaning that increasing the roughness, pitch and loudness of a bell sound leads to greater urgency. The mean urgency value pairs are presented using dark circles, while the smaller light dots depict the median values on the boxplot, accompanied by a best-fit line.

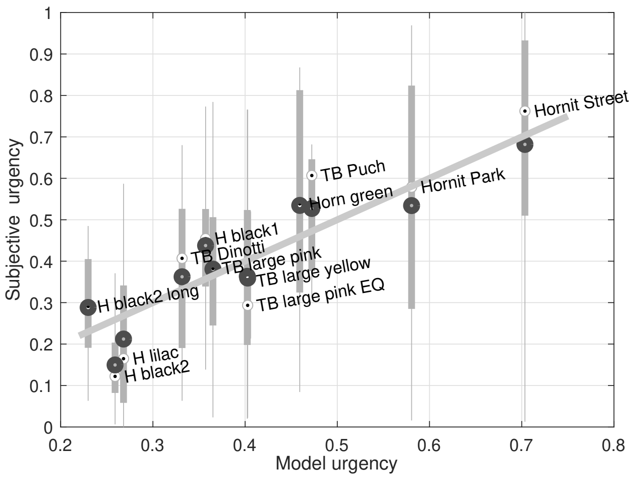



Figure 4: Boxplot of subjective urgency, plotted against objective model urgency. Circles correspond to mean (dark) and median (light). The model is shown as linear regression line.

$R^2$, the coefficient of multiple determination, provides a measure of how well the observed outcomes are simulated by the model. As there is a greater variance for urgency ratings between the 17 subjects for some stimuli than others, calculation of $R^2$ from raw data only yields a rather small value of 0.27. Therefore the corrected $R_c^2$ is used, which is referring to the mean of the subjective ratings, compensating the mentioned effect and leading to a value of 0.84.

## 6. GENERAL DISCUSSION AND CONCLUSION

We presented an evaluation of the perceived urgency and recognizability of commonly available bicycle bells in a silent environment as well as with background traffic noise.

A free identification task revealed that not all these warning signals are similarly identified as corresponding to a bicycle. While the twist type bell with the classic 'Rrring' sound achieved high recognizability values, hammer bells ('Ding') and large rotation bells ('Ding-dong') achieved lower but still high ratings. The toy horn, often found on children's bikes, yielded low recognition as bicycle, while an electric bike alarm of type Hornit was never identified as corresponding to a bicycle by the participants of the listening experiment.

An urgency rating experiment showed that the tested bell types differed significantly in urgency ratings. The highly irritating, siren-like electronic sounds of the Hornit bike bell got highest ratings for both no-noise and noise settings. However, in background noise not all test subjects were drawn to these sounds with the same strength as they were in silence.

The honk of the bike horn also achieved high urgency values for both traffic noise conditions, while average values were obtained by the group of rattling 'Rrring' sound bells. The most common bell types of group (2) 'Ding' and (1) 'Ding-dong' character find their place at the bottom of the urgency scale. Especially, in a noisy environment these sounds seem to lack the ability to stand out and attract attention.

No matter how high urgency scores were for each bell sound, the connection of recognizability and urgency must be considered. Bells which provide a good compromise between respectable urgency and high bicycle typicality, would be the classic 'Rrring' bells with the rattling character of group (3), demonstrating an even better performance under traffic noise conditions. As stated before, most people are not familiar with the Hornit sounds used as a bike bell, therefore these sounds do not appear suitable if one wants to be recognized as a cyclist. Hammer bells are definitely recognizable as bike bells, but their urgency is below average, in contrast to the bike horn with better urgency ratings but poor bicycle typicality scores.

Drawing conclusions from acoustical analysis of the bells' signal characteristics did not turn out to be done easily. Evaluating the chosen signal parameters of Sec. 4 did not always lead to explicit practical results, but it certainly does provide some insight.

Onset time and spectral centroid do not seem to play a major role in urgency prediction at all and perceived pitch seems to have only a little influence. Roughness appears to have an effect on urgency, but the sensory dissonance on its own can not explain the values of the test data entirely. A higher signal power surely influences the perception of urgency, even for our relatively balanced stimuli setup.

The simple objective urgency model showed a relationship for three of the parameters chosen: pitch, signal power, and roughness. In the case of the latter especially, the results support our assumption made in the experimental planning process, that the roughness of a sound has a positive effect on its perceived urgency.

For further investigation, other signal parameters need to be found and examined. Gaining satisfying results from observing only one parameter at a time is not possible for these already existing bell sounds. Many factors from spectral composition, temporal behavior, environmental noise level and, in particular, the acculturation and the listener's knowledge of meaning form a sound's perceptual urgency.

The sense of hearing will always remain important for the human being as a tool for localization and evaluation in possibly dangerous environments such as urban traffic. This is why research and new developments in the field of auditory warning design are of great value and importance. Improving the status quo should be encouraged and supported, thus guaranteeing a safer environment for everybody. Determining and grading warning signal urgency by spectral or temporal manipulation could be an effective approach leading to the desired objective of adequate warning sound design.

## 7. REFERENCES

[1] S. James, "Audio warnings for military aircraft," *Audio Effectiveness in Aviation*, 1997.

[2] J. Edworthy, S. Loxley, and I. Dennis, "Improving auditory warning design: Relationship between warning sound parameters and perceived urgency," *Human factors*, vol. 33, no. 2, pp. 205–231, 1991.

[3] R. D. Patterson, *Guidelines for auditory warning systems on civil aircraft*. Civil Aviation Authority, 1982.

[4] A. Guillaume, C. Drake, M. Rivenez, L. Pellieux, and V. Chastres, "Perception of urgency and alarm design." Georgia Institute of Technology, 2002.

[5] A. Pirhonen and K. Tuuri, "Communicative functions of sounds which we call alarms," in *Proceedings of the 16th International Conference on Auditory Display ICAD 2010*, 2010, pp. 279–286.

[6] A. Guillaume, "Intelligent auditory alarms," *The Sonification Handbook, Logos (Berlin, 2011)*, pp. 493–508, 2011.

[7] S. Langlois, C. Suied, T. Lagear, and A. Charbonneau, "Cross cultural study of auditory warnings." International Community for Auditory Display, 2008.

[8] J. Edworthy and N. Stanton, "A user-centred approach to the design and evaluation of auditory warning signals: 1. methodology," *Ergonomics*, vol. 38, no. 11, pp. 2262–2280, 1995.

[9] G. Lemaitre, P. Susini, S. Winsberg, S. McAdams, and B. Letinturier, "The sound quality of car horns: Designing new representative sounds," *Acta acustica united with Acustica*, vol. 95, no. 2, pp. 356–372, 2009.

[10] Federal Signal: Rumbler Siren, last access on 2018-05-03. [Online]. Available: https://www.fedsig.com/product/rumbler

[11] O. Pflüger, "Platz da! (engl. 'Make way!')," *brand eins*, vol. 19, pp. 14–15, 09 2017.

[12] "Ambulances to jam car radios in Sweden," BBC News, last access on 2018-05-03. [Online]. Available: http://www.bbc.com/news/technology-38650441

[13] A. Petocz, P. E. Keller, and C. J. Stevens, "Auditory warnings, signal-referent relations, and natural indicators: Rethinking theory and application." *Journal of experimental psychology: Applied*, vol. 14, no. 2, p. 165, 2008.

[14] The Hornit dB140, last access on 2018-05-03. [Online]. Available: https://www.thehornit.com

[15] W. A. Sethares, *Tuning, timbre, spectrum, scale*. Springer Science & Business Media, 2005, p. 32.

[16] O. Lartillot and P. Toiviainen, "A matlab toolbox for musical feature extraction from audio," in *International Conference on Digital Audio Effects*. Bordeaux, FR, 2007, pp. 237–244.