EUROPEAN COMMISSION
DG RTD

SEVENTH FRAMEWORK PROGRAMME
THEME 7
TRANSPORT - SST
SST.2011.RTD-1 GA No. 285095

eVADER

Electric Vehicle Alert for Detection and Emergency Response

Deliverable No.	D2.2	
Deliverable Title	Perceptual test 1: Detectability	
Dissemination level	Public/Confidential/Restricted	
Written By	Ryan L. Robart (INSA Lyon)	Nov. 5th, 2012
Checked by	Etienne Parizet (INSA Lyon)	Nov. 6th, 2012
Approved by	Juan Garcia (IDIADA)	Nov 6th, 2012
Issue date	Nov 6th, 2012	

Executive summary

The need for a solution to the increasing number of dangerously quiet cars as outlined deliverable 2.was the focus of the perceptual research conducted in work package 2.2. This research specifically focused on the detectability and localizability of experimental sounds synthesized according to sound features predicted to facilitate listener performance equitable to that of internal combustion cars. A primary concern for the consortium was to find a combination of sound features that may allow for such performance while maintaining sound levels that would not encroach upon the overall surrounding noise level of an urban soundscape. This research report summarizes the technical and research design used in our endeavor, as well as the human participant samples used in this set of studies.

Generally, the results confirmed that the 3 selected sound features; harmonic complexity, frequency modulation and amplitude modulation are all important for a suitable replacement sound. However, these features are not equally influential. In line with predictions, when amplitude modulation increased over 3 levels, listener performance increased as well. The influence of both frequency modulation, and harmonic complexity had an overall inverse effect on listener performance, which was not predicted. As a result of the limitations of a fractional design, interactions were not predicted, but did seem to play a role in variance that could not be accounted for.

Despite the limitations of a fractional design, the results were somewhat systematic. Results showed that 2 sounds were detected as quickly as the Diesel. Surprisingly, one sound produced half as many errors as the Diesel. These results have lead to clear recommendation for stimulus *313* for the prototype eVADER vehicle. Overall, it can be concluded that it is possible that a well-designed, quiet sound can equally, or even more effective as a Diesel engine which is ~ 10 dBA (peak level) louder in the virtual realm. These and other conclusions will be discussed along with suggestions for future research and potential risks.

Contents

1 Introduction

It has been established that quiet cars, such as hybrid and electric vehicles (*EV*), are potentially very dangerous for pedestrians. This is especially true for the visually impaired population. Taking this into account, we tested a diverse sample of participants including visually impaired persons as well as sighted persons of various ages from Germany, England and France. Five laboratories were used to conduct these experiments. They are as follows:

1) INSA (Lyon, France);
2) PSA (Vélizy, France);
3) Nissan (Sunderland, United Kingdom);
4) LMS (Leuven, Belgium);
5) TUD (Darmstadt, Germany).

At the suggestion of the European Blind Union (EBU), it was determined that increased noise caused by traffic and weather (e.g. rain) can combine for a most dangerous and confusing situation when a pedestrian relies on sound to navigate. In the interest of these concerns, it was decided to conduct most tests under these conditions. However, two labs (LMS & TUD) conducted tests using only traffic (no rain). These tests have not yet been completed and will be reviewed in an appendix. As a result this research report will focus only the tests (rain) conducted by INSA, PSA, and Nissan.

As outlined in the D2.1, it was predicted that 3 sound features were the best candidate features for the design of a sound that would ensure pedestrian safety while allowing for a lower level emission than the sound of an internal combustion engine. These 3 features are as follows:

1) Frequency modulation
2) Harmonic complexity
3) Amplitude modulation

Since these features were chosen based on a converging evidence review of the perceptual literature, it could only be predicted that listener performance (speed and accuracy) would increase with increasing levels of each factor. No predictions were made concerning the interactions of these factors. The following sections will review and summarize stimuli design, participant data, tasks, research design and results. The results will be discussed as well as risks and future experiments.

2 Stimuli and Soundscape Design

This section will briefly review the stimulus selection and design as described in the stimulus design proposal. Additional steps that were taken during the sound design process will be reviewed as well.

2.1 Waiting to Cross Scenario

As outlined in WP 1, several listening scenarios were recorded at the Idiada and Renault test tracks. Due to time constraints, it was decided that only the following scenario be tested in the experiments conducted in WP 2.2 (figure 1).

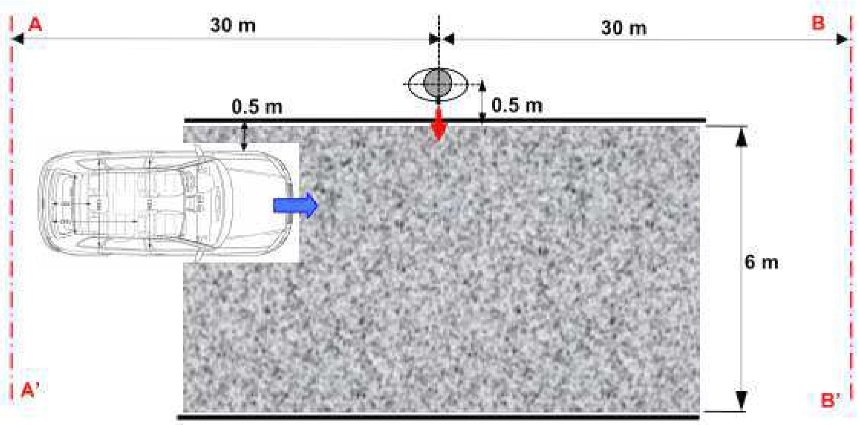


Figure 1: Graphical depiction of scenario 7 (see D1.6), or the ‘waiting to cross’ scenario. The recordings for the 20 kmh were used to make stimuli.

2.2 Sound Feature and Level Combinations

Recall that the 3 sound features selected were tonal complexity (number of harmonics), frequency modulation, and amplitude modulation (table A).

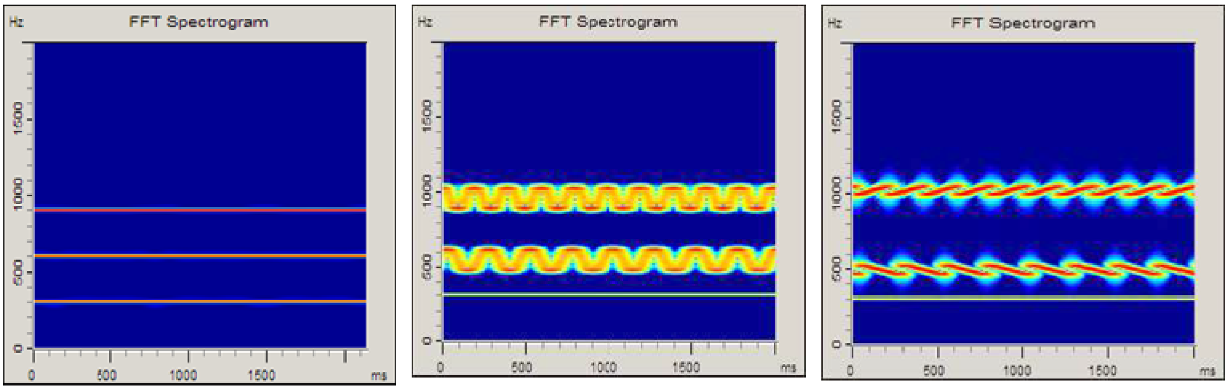
Stimulus	Frequency Mod (detuned)	Tonal Content	Amplitude Mod
1	Level 1	Level 1	Level 1
2	Level 1	Level 2	Level 2
3	Level 1	Level 3	Level 3
4	Level 2	Level 1	Level 2
5	Level 2	Level 2	Level 3
6	Level 2	Level 3	Level 1
7	Level 3	Level 1	Level 3
8	Level 3	Level 2	Level 1
9	Level 3	Level 3	Level 2

Table A: Taguchi table for fractional design for 9 stimuli with 3 levels each.

Since there was no precedent for sound design, the choices of frequencies and patterns of modulation were somewhat arbitrary. For this reason, these sounds were intentionally kept relatively simple.

Factor 1: Frequency Modulation

The frequency modulation patterns selected were as follows (figure 2).



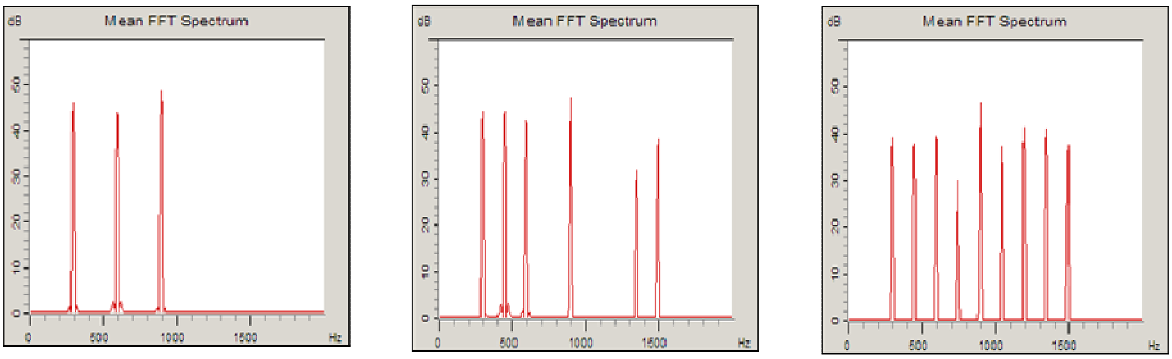
Level 1 : no modulation   Level 2 : sinusoidal modulation   Level 3 : saw-tooth modulation

Figure 2 : spectrogram of three sounds, according to the three levels of factor 1.

In the level 2 condition, the highest 2 frequencies oscillated between 5% above and below the harmonic. The frequency of these oscillations were designed to be enharmonic; 5 Hz for the highest harmonic and 4 Hz for the 2nd highest harmonic. This was done to avoid artifacts due to harmonic oscillations. The same design parameters used in level 2 were maintained in level 3 to ensure that any differences in responses would be due to the shape of the oscillation alone.

Factor 2: Complexity (Number of Harmonics)

The frequencies selected for the stimuli were based on the size of the loudspeakers (as outlined in D2.1) for the lower bound (300 Hz) as well as frequencies known to be annoying (see D2.1), and less useful for age related hearing loss in the upper bound (1500 Hz). The frequency distribution can be seen in the following figures (figure 3).



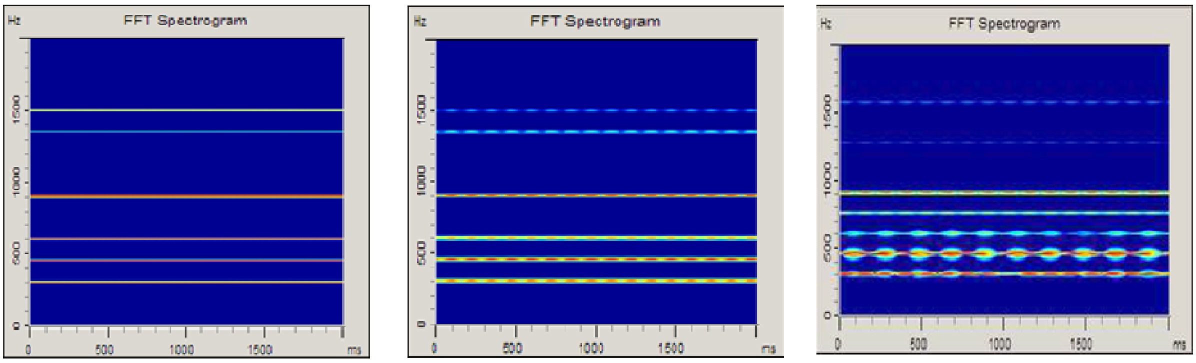
Level 1 : 3 harmonics   Level 2 : 6 harmonics   Level 3: 9 harmonics

Figure 3 : frequency spectrum of three sounds, according to the three levels of factor 2.

At level 1, sounds contained the three harmonics : 300, 600 and 900 Hz. At level 2, these harmonics were : 300, 450, 600, 900, 1350, and 1500 Hz, while they were 300, 450, 600, 750, 900, 1050, 1200, 1350, and 1500 Hz at level 3.

Factor 3: Amplitude Modulation

The amplitude modulation patterns selected were as follows (figure 4):



Level 1 : no modulation   Level 2 : sinusoidal modulation   Level 3 : irregular modulation

Figure 4 : spectrogram of three sounds, according to the three levels of factor 3.

At level 2, the amplitude envelope of all harmonics oscillated between < 20 dBA to maximum amplitude at 8 Hz. All frequencies in these sounds were modulated according to the same envelope. At level 3, there were 4 separate amplitude envelopes used in these stimuli in order to create an overall sound that had time-varying structure. It should be noted that one of the envelopes was the same as the level 2 envelope (8 Hz) to maintain some continuity between levels. A detailed description can be found in Annex 1.

2.3 Stimuli Synthesis and Recording Processes

Various labs were involved in the recording and synthesis process as shown in the figure below.

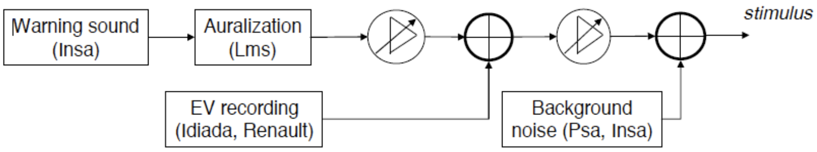


Figure 5: Path diagram showing the different lab contributions to sound synthesis and recording (see D1.6).

All recordings were made with a dummy head (Head Acoustics HMS III). As described in D1.6 the peak level of the internal combustion vehicle (diesel) was approximately 76 dB(A), while the peak level for the electric vehicle was measured at approximately 70 dB(A).

As previously described, 9 sounds were designed according to a fractional Taguchi matrix. These 9 sounds were synthesized and recorded with a custom synthesizer (eVADER Synth) using Max/Msp. The 9 sounds were level equalized (in dBA) by using Matlab at INSA. These sounds were then further synthesized by LMS (Matlab) using algorithms that modeled a sound source moving at (20 km/h) on a textured semi-reflective surface such as concrete on a street, as heard by a pedestrian facing the road (using head related transfer functions). All the sounds were 10.8 seconds in length, in accordance with the vehicle recordings. Recordings were passed through an inverse filter designed in Matlab to correct for the frequency response of the headphones used in the experiment. Once the sounds were modeled by LMS, they were layered onto the recordings of the *EV* by INSA. The result was 9 new stimuli, each composed of 1 synthesized sound and the *EV* recording.

The relative levels of warning sounds and background noise was adjusted by a trial-and-error process, in order to fulfill the following rules:

- The detection of the electric vehicle should be rather difficult, while the detection of the diesel car should be easier.
- Adding a warning sound should not make the detection too easy.

All 11 sounds were then channel swapped, so that both possible directions (left->right and right->left) of pass-by could be heard. As a result, 22 stimuli were designed. The levels dB(A) of all stimuli were recorded and measured by tracking the stimuli through the Stax headphones (placed on the dummy-head), which was ported via XLR into a computer. This was done to ensure that original levels of the vehicle recordings were not significantly changed (figure 6).

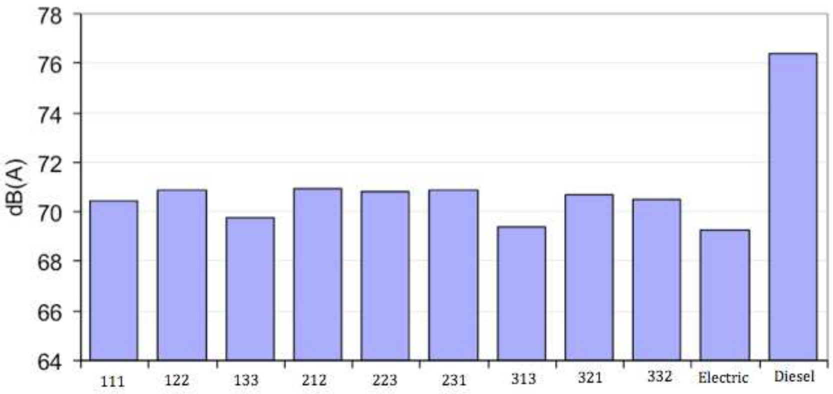


Figure 6: Peak level (dB(A)) of sound stimuli.

Figure 6 shows that the warning sounds increased the peak level of the electric vehicle by 2 dB(A) only, so that peak levels of all electric vehicles is much lower than the one of the Diesel car (the difference is more than 5 dB).

2.4 Soundscape (Background Sound) Design

In a detection task protocol, adding a background noise is necessary. In the interest of maintaining a high level of ecological validity, it was decided to use a continuous and natural background sound. This background noise was made of traffic noise, but rain noise was also used as this represents a challenging condition for visually impaired pedestrians, as suggested by the European Blind Union. This was done so that the soundscape not only sounded natural, but also would be challenging.

Traffic

Regarding the traffic, several recordings, made at various locations by Idiada and Renault, were auditioned. A recording of a busy auto-route in Velizy, France was selected because of its consistent, high-density traffic flow. This recording was thought of as a somewhat stationary (unchanging) background, as there few pauses in traffic flow, and relatively few loud and abrupt sounds. A 2 minute sample was selected that would be looped continuously during the experiment. In the interest of reducing potential confounds associated with binaural cues, only one channel of the selected sample was used. This channel was then divided into 2 channels, panned approximately 45 (right channel) and -45 (left channel). The background sound was then low-pass filtered and reduced in level to approximately 69 (mean level dBA) to emulate a busy roadway approximately 100-200 meters in front of the listener.

Weather (Rain)

A recording of rain was mixed with the continuous loop of traffic to create the completed background sound. As with the traffic sound, a single channel of the recording was made into stereo and panned to emulate natural sounding space. The overall level of the rain was then adjusted so that the completed background sound maintained an overall mean level of 69 dB(A).

This "wet" background noise was used at INSA, Nissan, PSA and TUD, while a "dry" background noise was used at LMS and TUD. In the following, the analysis will be conducted on data collected with the "wet" background noise.

3 Design and Materials

As outlined in the D2.1, a 3 (sound feature) X 3 (level) fractional repeated measures design was used to measure listener detection and localization of the recordings of approaching cars (stimuli). The unmodified recordings of the *EV* and the *Diesel* served as anchors. In other words, the *Diesel* was expected to facilitate the fastest and most accurate responses, while the *EV* was expected to produce the slowest and least accurate responses. Recall that the 9 stimuli containing the synthesized sounds were designed according a Taguchi matrix for fractional designs (table B).

Stimulus	Frequency Mod (detuned)	Tonal Content	Amplitude Mod
1	Level 1	Level 1	Level 1
2	Level 1	Level 2	Level 2
3	Level 1	Level 3	Level 3
4	Level 2	Level 1	Level 2
5	Level 2	Level 2	Level 3
6	Level 2	Level 3	Level 1
7	Level 3	Level 1	Level 3
8	Level 3	Level 2	Level 1
9	Level 3	Level 3	Level 2

	Anchors (Control Stimuli)
10	Electric Vehicle (EV)
11	Internal Combustion Engine (Diesel)

Table B: Taguchi table used for the experiments design (top portion), and the two anchors predicted to produce the top and bottom boundaries of performance (bottom portion).

After swapping the right and left channels of each of these, there were a total of 22 stimuli. Each stimulus was presented 4 times over 88 trials. The experiment was divided into 2 blocks and each stimulus was heard twice per block. Trials were randomized separately for each participant. The total duration of a typical experiment was approximately 45 minutes. In order to create a realistic listening experience, stimulus onsets were somewhat unpredictable as it is on a real street. A pseudo-random 1-20 second inter trial interval (ITI) was used for each trial. A new ITI was generated after each trial to achieve this realistic timing between approaching cars (stimuli). Trial onsets were concurrent with stimulus onsets. There was no limit for number of responses, and responses had no impact on the stimulus presentation. For example, if a participant pushed the ‘left’ button in response to an approaching stimulus, the stimulus would continue to play regardless of the response. Keystrokes were recorded, coded and time stamped by the experiment software so accuracy and reaction time could be measured. No feedback was given to the subject.

The experiment was conducted in a dimly lit sound-attenuating chamber. A PC computer was used to present stimuli and record responses. Participants responded to stimuli using a standard keyboard. A Gina sound card was used for audio output. Stax headphones (Lambda Pro: electrostatic) and amplifier system was used to deliver the sound stimuli. The experiment was programmed using Delphi software for PC computers running Windows 7 operating system.

Participants were familiarized with all sounds with short demonstration during the instructions. Participants were informed that their task would be to listen to these recordings of approaching cars and respond as quickly and accurately as possible by pressing a computer key that corresponds to the direction of approach. After a participant completed a short training session (5 trials), they could begin the experiment. After the completion of the first block (44 trials), participants were given a short break. At the completion of the 2nd block, participants were debriefed and thanked for their participation.

120 participants (aged 20-72) either volunteered or were compensated for their participation. 84 participants had normal or corrected vision and 36 were visually impaired (VI). All the VI participants were blind, and nearly all were VI from birth or early childhood.

Most sighted and VI participants reported normal hearing. However, reports of hearing loss were common among participants over the age of 60. At Insa, hearing ability of subjects was measured before the experiment using an automatic audiometer device (AudioConsole).

4 Results

4.1 Detection

Participant data was rejected (8 participants) for analyses if they missed and/or made errors on more than 40% of the stimuli, or if a participant missed all of a certain stimulus. Thus the analysis was conducted on data from 112 participants (79 sighted and 33 VI ones).

Between-subject variance was fairly high. The averaged response time was computed for each subject. Values are between 1.3 and 5.2 seconds, which represents a distance to the car between 22.4 and 0.9 m. (figure 7).

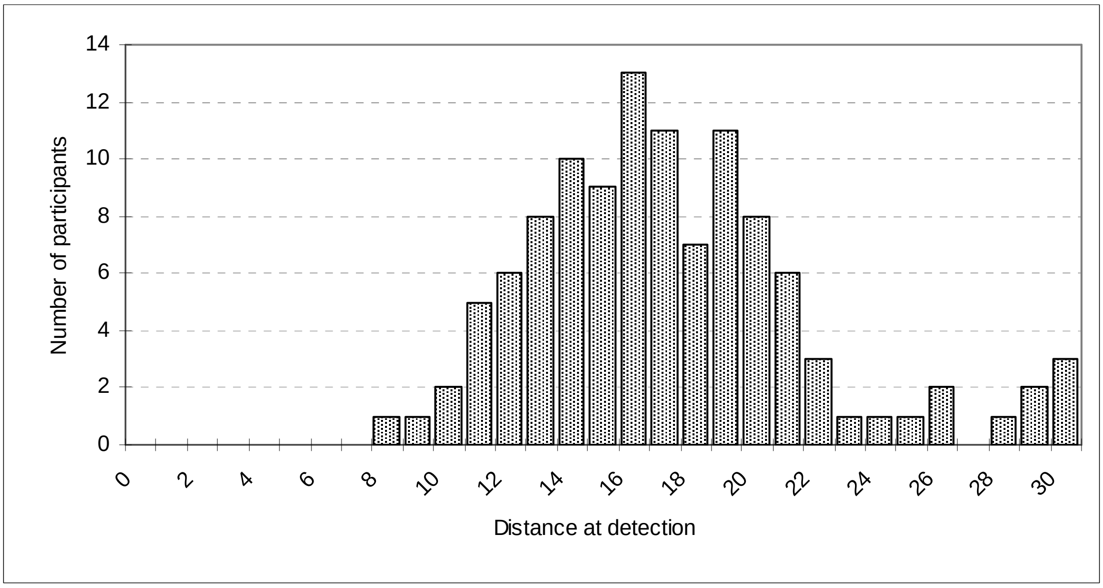


Figure 7: Repartition of individual averaged detection times.

These differences were not found to be systematic according to age or audio metric sensitivity. It was determined that the between subject variability must be due to cognitive factors, such as strategy. In cases like this, it is possible to minimize between subject variability by removing the amount of variance associated with each participant. This is done by centering the data by removing the difference between mean for each participant and the grand mean.

No differences were found between the first and second blocks, and there was not an effect of direction.

Furthermore there was no difference in average performance between sighted and VI participants (figure 8). A Kruskal-Wallis test indicates that the difference is not significant (*p* = 0.12).

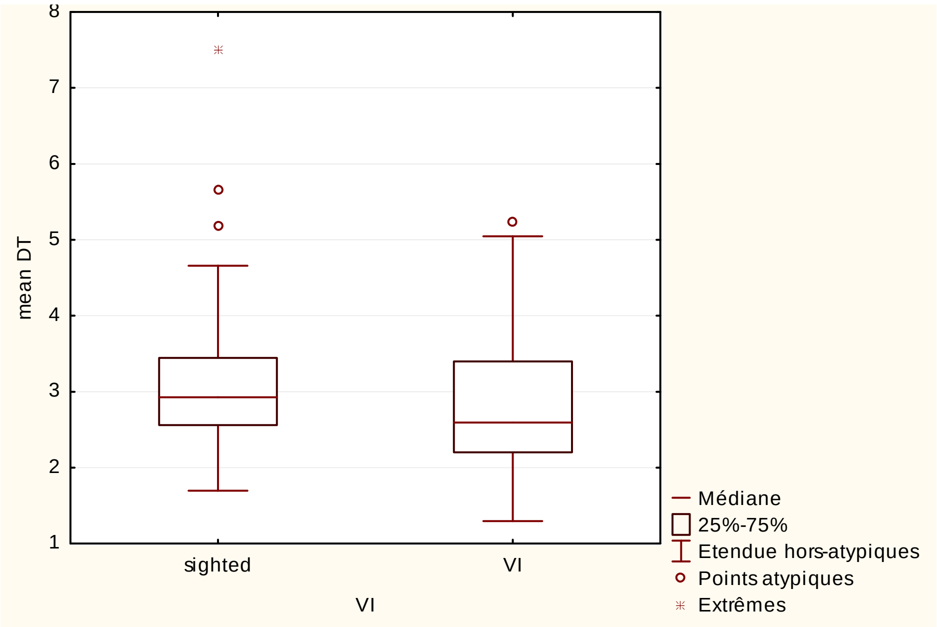


Figure 8 : boxplots of mean reaction time in seconds for sighted and VI participants.

Since there were negligible differences between blocks and direction, the data was collapsed across stimulus giving a total of 11 stimuli with 8 repetitions of each. Overall, the pattern of data was robust between different laboratories (figure 9).

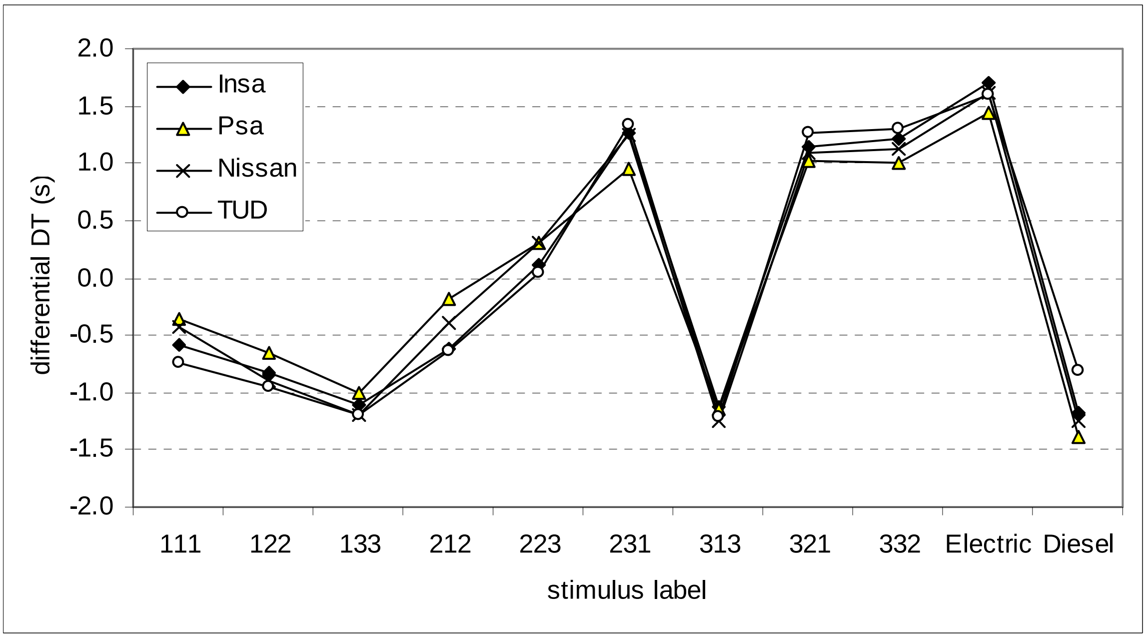


Figure 9: Lab comparison for reaction times after data was centered and collapsed across direction.

The reaction time data can be converted into a distance metric (meters) to achieve a more clear visualization of where the cars were when they were detected (see figure 10).

Figure 10: Conversion of reaction time to distance (meters) of all participants. The pink area indicates the danger zone as outlined in D1.1.

As expected, the electric vehicle was not detected at a safe distance (5-6 meters from the listener). Furthermore, reaction times for the diesel were in line with predictions, as it was heard 22 meters before crossing. The variation between the synthesized sounds is quite surprising. Three synthesized sounds: *332*; *321*; *231*; were all detected at unsafe distances. Interestingly, 2 of the synthesized sounds, *313* and *133* were heard at the same distance as the Diesel. The 4 remaining synthesized sounds: *111*, *122*, *212*, and *223* were also heard at safe distances. However, post-hoc t-tests showed that the reaction times to these sounds were different from the safest sounds: *313*, *133* and *Diesel*.

Omnibus ANOVA

A 3 (factor) X 3 (level) fractional ANOVA conducted on detection (reaction times) of all participants showed that all there were main effects for all 3 factors (table C).

	Sum of squares	d.o.f	Mean of squares	F	p
F1	319 532 000	2	159 766 000	601.5	<0.001
F2	255 117 600	2	127 558 800	480.3	<0.001
F3	296 256 300	2	148 128 100	557.8	<0.001
residu	265 854 600	1001	265 589		

Table C: ANOVA table showing the main effect for the factors *frequency modulation, complexity,* and *amplitude modulation.*

Certainly it can be said that all three factors are very effective at different levels (e.g. 1-2-3). This is the reason why the *F*-values are so high. Based on these results, it seems that all three factors have a similar impact on listener detection.

Pairwise comparisons conducted via 3(factor) X 3(level) X 2(interaction) ANOVA's suggest that the most powerful interaction occurred between factor 1 (frequency modulation) and factor 2 (harmonic complexity). Amplitude modulation was the only factor that did not produce significant interaction effects. However, interpretations of interactions are very limited in a fractional design because all possible combinations are not represented. However, the overall effect of each factor can be derived by center-reduction of the *F*-values (figure 11).

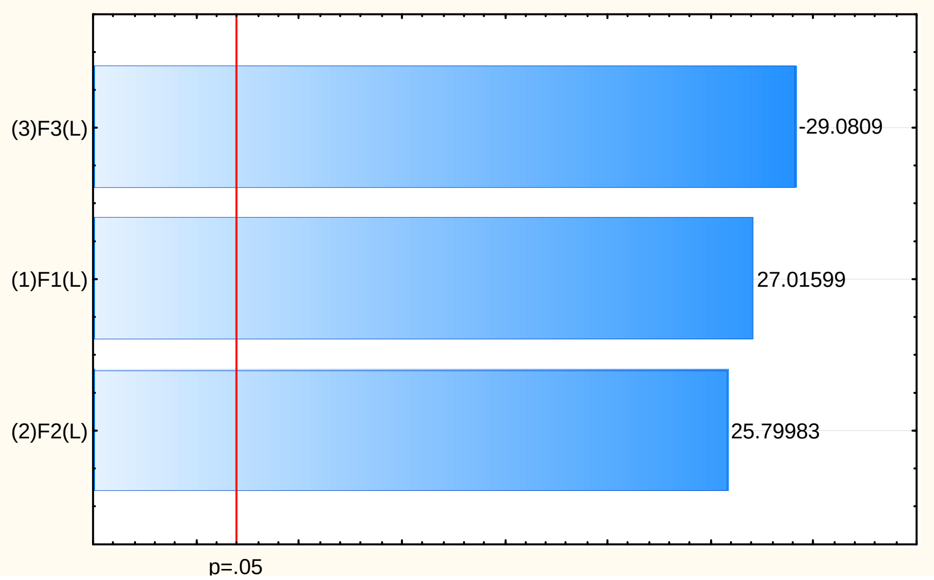


Figure 11: Factor effects figure demonstrating the relative contribution to main effects by all factors.

Amplitude modulation appears to have the largest effect, followed by frequency modulation and tonal complexity respectively. Still, the strength of the factors are comparable. However, the negative value associated with amplitude modulation indicates that as the level (1-3) of amplitude modulation increased, reaction times decreased (figure 20). This confirmed our prediction regarding the trajectory of the amplitude modulation effect. Conversely, reaction times were found to increase with the levels of both frequency modulation and tonal complexity (figure 12).

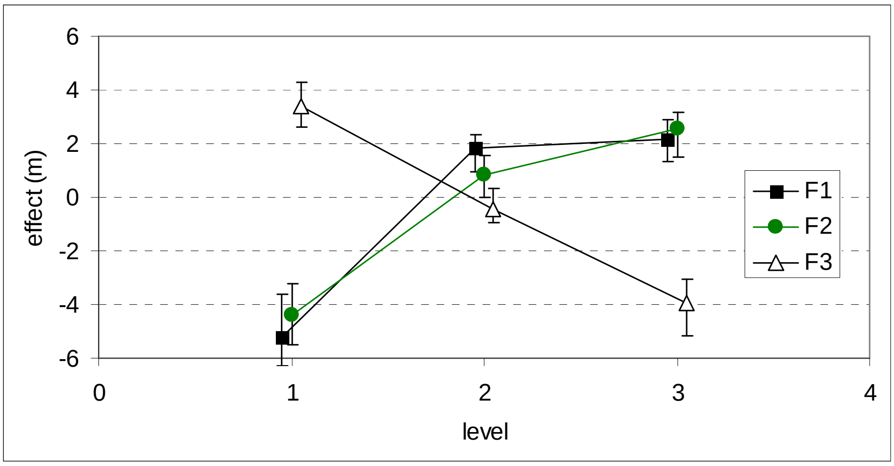


Figure 12: Effects of factors levels on listener performance. F1= *frequency modulation,* F2= *harmonic complexity,* F3= *amplitude modulation.*

One possible way of evaluating the validity of the model is to predict the measured values by using the formula :

$$T_r(i,j,k) = \overline{T_r} + E_{1,i} + E_{2,j} + E_{3,k} \tag{0.1}$$

Where : *i*, *j* and *k* are the levels of factors 1, 2 and 3;

$\overline{T_r}$ is the mean reaction time;

$E_{1,i}$ is the effect of factor 1 at its level i (and the same for $E_{2,j}$ and $E_{3,k}$)

The comparison of predicted and measured values (figure 13) shows that the model is not fully predictable. This indicates that some slight interactions should exist, as already mentioned.

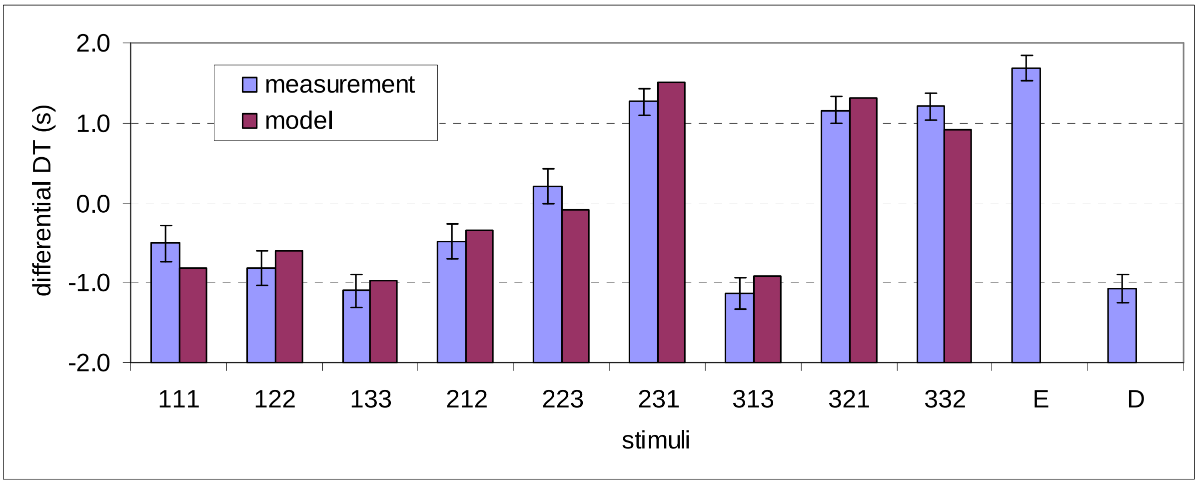


Figure 13: Model and data comparison for all participant detection times (centered).

As it can be seen in the chart above (figure 13), there is a relatively good fit. However, notable discrepencies exist for stimuli *122* and *332.* In that, the shape of the data distribution varies from the shape of the model distribution only in those areas. Based on the fact that the model assumes independent effects from the factors, it is likely that the variation between the model and data is due to some interaction effect.

4.2 Accuracy

Erroneous responses occurred when a participant responded incorrectly regarding the direction of approach. It was expected that there would be few errors overall. Indeed, the overall number of errors is 351, which is small when compared to the number of trials (112x88 = 9856.
Interestingly, the distribution of errors closely resembles the distribution shape of the reaction time data (figures 14 and 15).

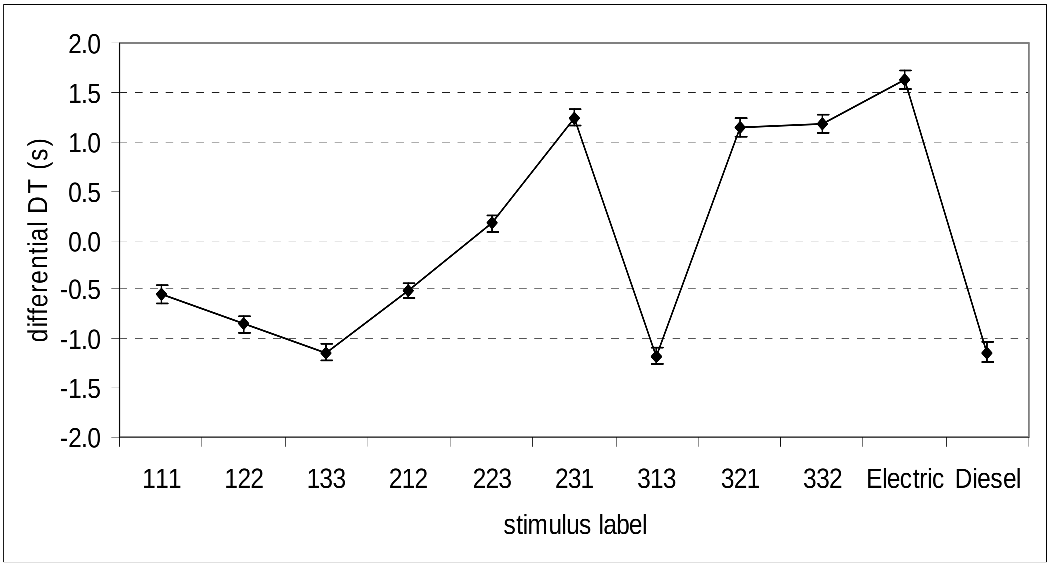


Figure 14: Average reaction times (centered) for all participants.

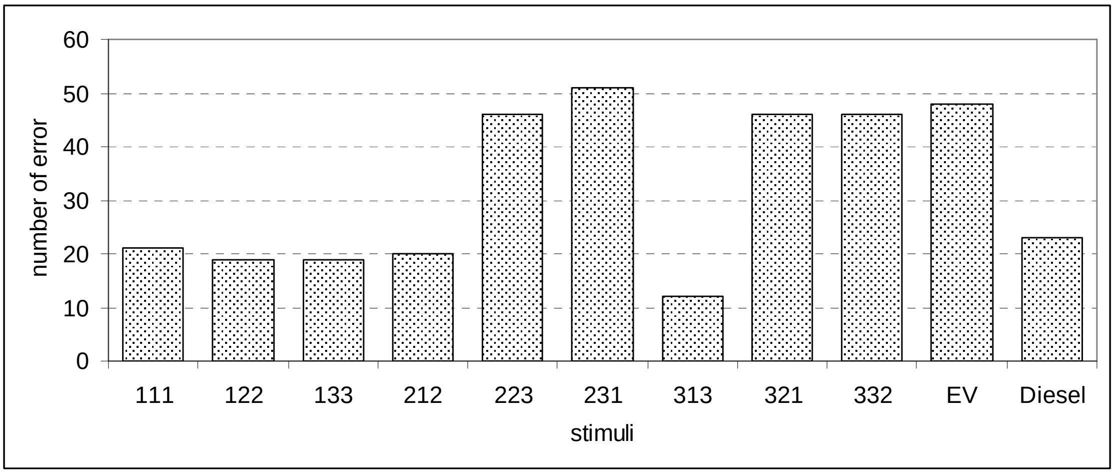


Figure 15: Number of errors for all participants for each stimulus.

Errors increased and decreased with reaction times, such that stimuli heard later produced the most errors and vice versa. However, it should be noted that the stimulus that produced the fewest errors was sound *313*. In fact, *313* produced only half as many errors as the other 2 safe sounds, *133* & *diesel*.

Is louder better?

These results cannot be the result of differences in sound pressure levels (see figure). There is clearly large difference in peak level (~ 5 dB(A)) between the 3 safe sounds. In fact, the 2 quietest (peak value) synthesized sounds were detected the earliest and most accurately (figure 16).

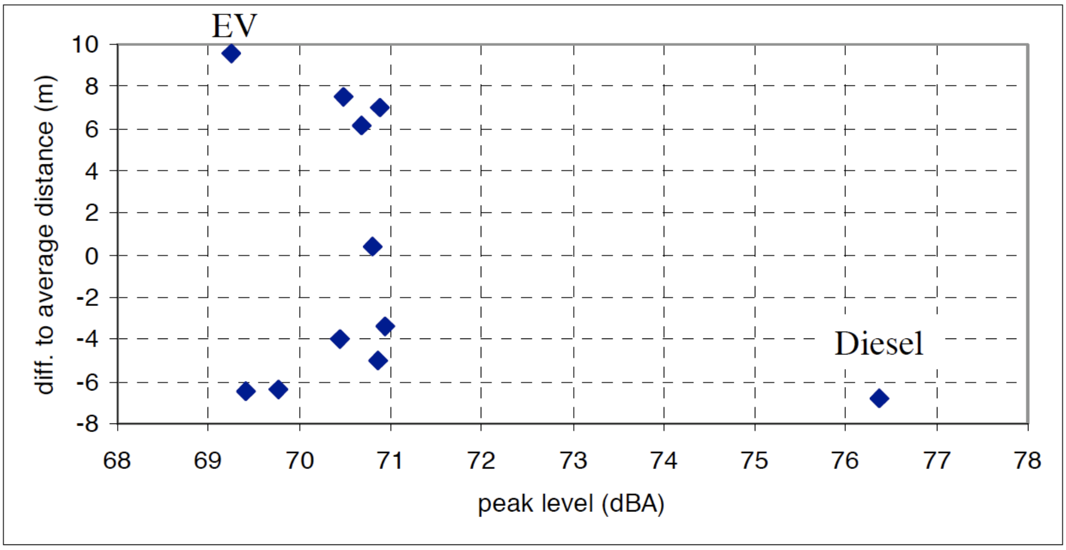


Figure 16: Comparison of average distance at detection, and peak level (dBA), for all stimuli. Note that the distance is independent of peak level (dBA).

5 Conclusions, Recommendations and Discussion

Taken together, the detection and accuracy results point to *313* as being the safest stimulus in our test, which was not expected. This was true for all kind of participants (sighted and visually impaired). Therefore, it is recommended that the safest sound, *313*, be utilized for the prototype eVADER vehicle. Based on this research, it is likely that sounds designed using the following featural constraints could be the safest and most effective:

1) Little or no frequency modulation (level 1)
2) Harmonic structure containing less than 6 harmonics (level 1)
3) Temporally variable amplitude modulation (level 3)

Unfortunately, the *113* combination was not required by the Taguchi matrix for fractional designs. Therefore, a sound containing this combination was not used in our experiments. Still, the results allow for models and predictions regarding the untested combinations. Models predict that '113' should produce the best results. However, it seems clear that *313* far exceeded expectations regarding detection and accuracy hence its recommendation.

Recall that the above *113* combination was not the pattern predicted to be optimal. Even though a pattern was not explicitly predicted to be optimal, it was expected that the optimal pattern would have been *313*. Basically, the prediction was that complexity would facilitate good listener performance. This was apparently true for amplitude modulation, be the results are quite the opposite for frequency modulation and harmonic complexity. This can be explained as the direct result of the low-level emission (dBA). Recall that all synthesized sounds were normalized to be the same average level (~ 65 mean spl dBA). As can be seen in the graph, the average reaction time decreased with the number of harmonics. This is likely because the spectral energy is more focused when there are fewer harmonics. So, if a sound has 3 harmonics and is the same overall level as another sound with 9 harmonics, the level of the 3 harmonics far exceeds the level of any harmonics in the 9 harmonic stimuli. This could be because the energy is spread out over more frequency bands in the 9 harmonic stimuli. Basically, it seems that 'less-is-more' when it comes to the frequency content of a quiet alarm sound.

Similarly, the less-is-more principle can help to explain why reaction times decreased with frequency modulation. It could be that the fluctuation of the energy in the high harmonics may cause a blurring of the energy in those bands, thereby spreading the energy and de-focusing the bands. Furthermore, as we know from Bregman's (1990) theory of stream segregation (also see the stimulus design proposal), enharmonic frequencies are not readily fused with harmonic frequencies. This could cause the auditory perceptual system to struggle to resolve the auditory image, vis a vis facilitating slower reaction times.

As far as amplitude modulation, the less-is-more principle does not seem to apply. Certainly, the effect of amplitude modulation is the most clear as it is almost perfectly linear. Even though it is tempting to suggest that amplitude modulation was the most influential factor, it may not be the case. If this statement were true the reaction times and accuracy for *113*, *313*, and *223* would be more congruent. Recall that the data for *113* and *313* were indeed quite congruent, but reaction times were much slower, and there were many more errors for *223* If temporally irregular amplitude modulation were the most important factor, it would be expected to overcome the other factors to produce results similar to its sibling stimuli. A potential explanation for this can be derived from analyzing the spectral content of '#2#' sounds which were among the worst stimuli (figure 17).

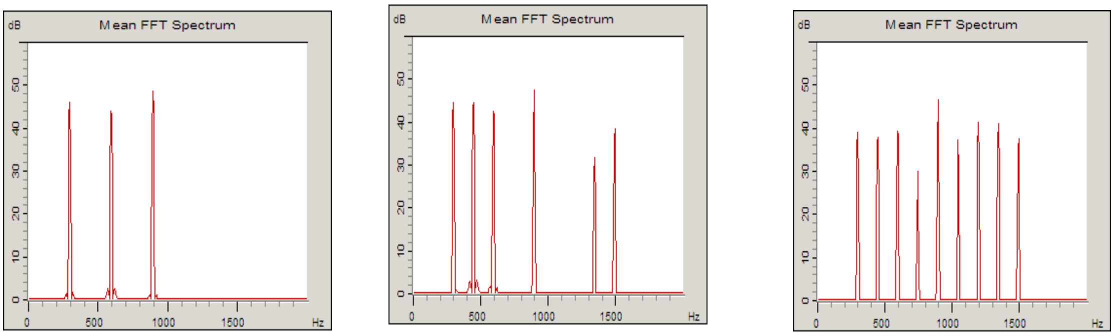

Level 1: 3 harmonics
Level 2: 6 harmonics
Level 3: 9 harmonics

Figure 17: Comparison of the harmonic structure of each level of factor 2 (*complexity*). The numbers below each image are in Hz.

As you can see, the 3 harmonic stimuli were composed of 3, 300 Hz steps. Similarly, the 9 harmonic stimuli were composed of 9, 150 Hz steps. Notably, the spectral content of the 6 harmonic stimuli may be considered 'less harmonic' than its sibling stimuli. There are obvious breaks in the harmonic path. The largest being a 450 Hz step from 900-1350 Hz. Essentially, the stepwise progression was not continuous as it was in the 3 and 9 harmonic stimuli. The auditory streaming literature would predict poor or slow fusion of these harmonics by listeners, which would be magnified if the higher frequencies were modulated. It would be interesting to design new 6 harmonic sounds that could be 'more harmonic' to test if it would increase the effect of amplitude modulation. Certainly, this question should be tested in the future. Still, this demonstrates the lack of a dominance of any of the 3 features used to design the stimuli.

In the same vein, the reader is reminded that the temporally irregular envelope structure, which arguably produced the best results, was arbitrarily designed. There are innumerable possibilities for such a pattern, and it is highly probable that there are upper and lower limits to the various parameters used in such patterns (e.g. speed and regularity). For this reason, it is likely that any candidate pattern for a given manufacturer must be rigorously tested in similar experiments to ensure it can produce listener performance similar to what we found here. Perhaps the greatest limitation of this experiment is that only one such temporally irregular amplitude modulation envelope was tested. This fact alone necessitates trepidation when drawing conclusions.
Still, according to the results, it can be cautiously concluded, that a well-designed sound can produce early and accurate detection at levels (dBA) that are not louder than the EV alone. Certainly though, more research is needed before any strong theoretical claims can be made. In that, we cannot confidently conclude explain how frequency modulation, tonal complexity, and amplitude modulation interact to produce our results.

Future research should include all combinations of features to obtain a design where interactions can be explored more completely. Furthermore, it is imperative to conduct further tests with different recorded scenarios in the interest of pedestrian safety (see WP1). For example, based on the research presented here, it cannot be determined if a sound like *313* is effected by the listener position with respect to the vehicle. Moreover, it is still unknown what the effect of multiple, similar sound sources might have on listener performance. For example, more often than not, there are multiple cars on a street. Based on the masking literature, it can be assumed that multiple, similar sound sources (e.g. 3 *313s*) could create a confusing and potentially dangerous situation for a pedestrian as a result of masking. However, it might be most important to explore if vehicle dynamics like acceleration could be paired to the synthesized sounds so that certain features change according to what the vehicle is doing (e.g. decelerating or moving slower). This may provide for a release from masking, but this question can only be answered empirically. The aim of the next project ‘sound meaning’ will be to answer such questions.

Annex 1: Sound *313* Synthesis

Frequency Modulation (table D)

frequencies	600 Hz	900 Hz
Frequency modulation	saw-tooth	saw-tooth
amount (Hz)	150	150
range	525-675	975-825
frequency (Hz)	4	5

Amplitude Modulation Envelope Parameters (table E)

Parameters	Amplitude Modulation			
envelope	1	2	3	4
dc offset	0.51	0.7	1	Ring
amplitude	100%	100%	100%	900%
frequency of modulation	8 Hz	3 Hz	33 Hz	5 Hz

Envelope 4 was compressed to maintain as much of the steeper slope as possible without dominating the over-all sound pressure level of the entire sound.

There were 4 distinct amplitude envelopes, and each frequency band was assigned to one master envelope. As a result, the lowest frequency (300 Hz) was modulated in the same way as all frequencies in the Level 2 amplitude modulation category (envelope 1) The 2 higher frequencies had the same m)aster envelope (envelope 4), which is known as a ring-modulation. Ring modulation amplifies all frequency components equally, which means the DC offset = 0, or 0%. This means that the carrier frequency is inaudible. More importantly, it means that the signal is dipolar, making the slope of the amplitude function steeper. It is important to note that envelopes 1 and 2 have carrier frequencies that are unlikely to be audible. However, envelope 3 was given an enharmonic and potentially audible carrier frequency that would cause the envelope to change over time. Still, the reader is reminded that envelopes 2 and 3 were not assigned to any frequency as master envelopes. A master envelope is the amplitude envelope by which a given frequency was constantly modulated (figure 28).

Master Amplitude Modulation Envelopes (table F)

Frequencies	Master Envelope
300	1
600	4
900	4

All 4 envelopes were assigned to be a sub-envelope to at least one frequency (table F). This means that while a given frequency is always modulated according to its corresponding master envelope, the master-modulated sound was periodically modulated by a sub-envelope amplitude modulation. In order to achieve a temporally irregular amplitude modulation structure, the master-modulated sounds were sub-modulated with a sequential structure. This means that the master-modulated sounds were sequentially cross-faded (panned) through the assigned sequence of sub-envelopes.

Sequential Envelope Structure (table G)

Frequencies	Sequential Envelope Structure (sub-envelopes)		
300	2	3	1
600	2	4	
900	1	1	

In order to ensure that there would be unexpected transitions in the sequences, dynamic panning parameters were enforced on the cross-fades.

Dynamic Panning Parameters for Irregularity (tables H-J)

Dynamic Panning Parameters (300 Hz)	
master envelope	1
sub-envelopes	2--3--1
pan 1	time-synced
time delay	1066 890
pan 2	smooth
time speed	1066 300

Dynamic Panning Parameters (600 Hz)	
master envelope	4
sub-envelopes	2--4
pan	time-synced
time delay	1036 200

Dynamic Panning Parameters (900 Hz)	
master envelope	4
sub-envelopes	1--1
pan 1	time-synced
time delay	1080 100
pan2	time-synced
time delay	1080 12

The separate frequencies had distinct dynamic panning parameters. Recall that each frequency was assigned a master envelope. The master-modulated sound was then cross-faded (panned) through a series of sub-envelopes according to these dynamic panning parameters. There were 2 possible panning assignments for each frequency.   The first pan (pan1), controlled the amount of the master-modulated frequencies sent to the first pair of sub-envelopes in the sequences. The second pan (pan 2) controlled the amount of the master-modulated frequencies sent to the 2nd pair of sub-envelopes from the 2nd of the first pair of sub-envelopes. So, even though 600 Hz and 900 Hz had only 2 sub-envelopes, dynamic panning still occurred between those sub-envelopes and their corresponding master-envelopes.

As you can see in the figures above, a given pan (1 or 2) can be time-synced or smooth. If a panning function were time-synced, it would transition from one sub-envelope directly to another. On the other hand, if the panning function was smooth, the master-modulated sound would smoothly be cross-faded from one sub-envelope to another depending on the sequence. A separate metronome was used to control the timing of each panning function. The time-value in the charts above illustrates the time between panning function onsets. The value associated with the delay parameter refers to the time (milliseconds) that a the panning function delays the onset of the pan (cross-fade) to the next stage in the sequence. The value associated with the speed parameter refers to the amount of time (milliseconds) that a smooth panning function cross-fades from one channel to another.

As you can see in the figures below (18, 19), there is little regularity in the overall temporal nature of the amplitude modulation structure using these parameters outlined above. When the *313* waveform is visually compared to a *312* waveform the differences in regularity are obvious.

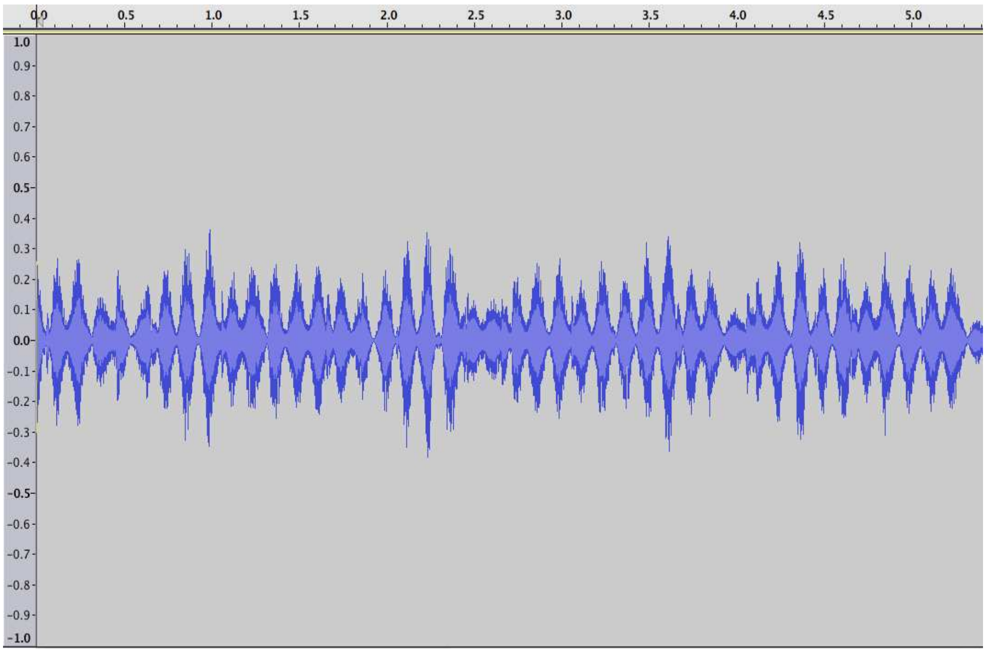


Figure 18: Graphical representation of *313* over 5.4 seconds, or 30 meters.

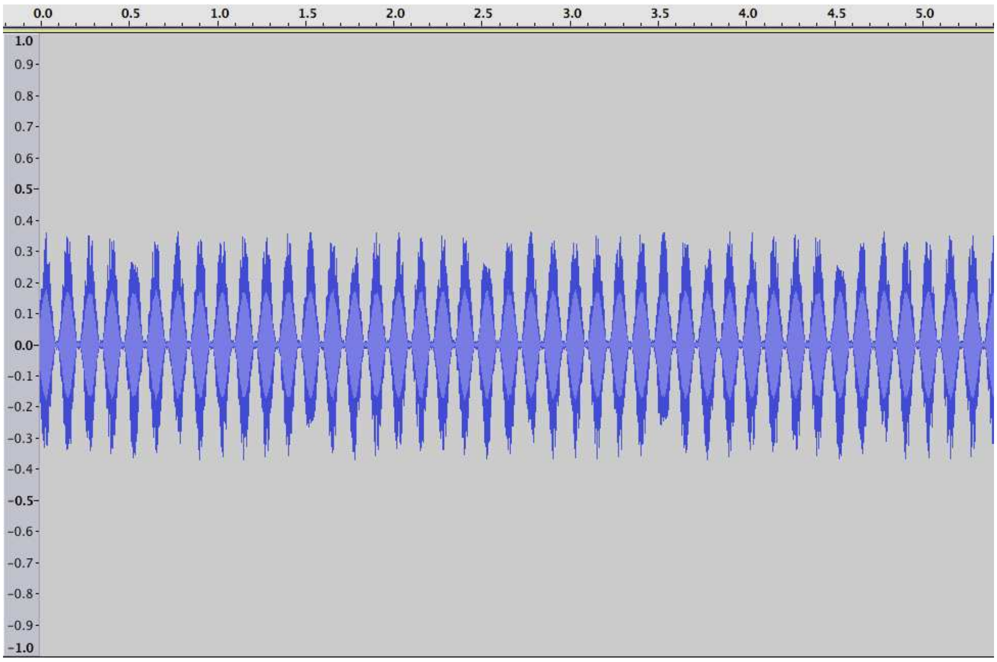


Figure 19: Graphical representation of *312* over 5.4 seconds, or 30 meters.

Annex 2 : Test instructions

Informed Consent:

This experiment concerns auditory perception. You will be presented auditory stimuli and you will be compensated for your participation.

Your data will be held strictly confidential. You will be allowed to ask questions at any time, an will be given periodic breaks throughout the session. You have the right to end your participation at any time. However, if you should choose to end the experiment early, any monetary compensation will be withheld. Please feel free to contact (name of principal investigator at laboratory) with any questions regarding this research.

Do you have any questions regarding these issues before providing your consent to participate in this experiment?

If I have your consent, there are a few more questions I must ask you before we begin.

Pre-Experiment Questions:

- What is your age?
- Do you have a dominant hand or are you ambidextrous?
- Do you have normal vision (sighted participants)?
- If you have corrected vision, can you tell me your prescription (sighted participants)?
- Do you have normal hearing?
- Do you have any other physiological decrement that may impinge on your ability to safely walk without aid?

***Participants will verbally provide information regarding their age and any relevant long term physiological decrements (e.g. myopia, amblyopia, congenital blindness, tinitus etc.). All decrements must be noted by the experimenter. The experimenter should explicitly state that all personal information divulged by the participants will be kept confidential. The experimenter should note any aid that participants divulge. For example, many visually impaired people use canes or dogs. Also, if a participant uses a hearing aid but is not deaf without it, we will want to know this.*

Thank you for that information.

Are you ready to begin?

Procedural Instructions:

Summary:

Background: You may have noticed in your everyday life that cars are increasingly becoming less noisy. You may have had an experience while walking where you were surprised by a very quiet car. Recent research suggests that quiet cars such as electric and hybrid vehicles may be dangerous to pedestrians, especially in a noisy environment.

Goal: We are interested in learning about the important aspects of the sound that pedestrians may use to avoid being struck by such cars. In our experiment, we have added synthetic sounds to some recordings of some cars while adding no sound to others. These recordings will be presented within the context of a realistic noisy urban environment. The noise consists of heavy traffic on a distant freeway on a rainy day.

Task: Your task will be to detect and localize only the cars that are approaching and eventually passing by on a street directly in front of you.

Detailed environmental description: We ask that you imagine that you wish to cross the street, and that you are facing perpendicular to the suburban street waiting for a safe time cross. From time to time you will hear various cars approaching and crossing directly in front of you from either the left or the right. There is not a traffic light, so the cars will not slow down, or stop. The speed and direction of these cars were controlled by us and were deliberately recorded for the purposes of this experiment. So, in sum:

1. you imagine that you are within the recorded environment that you hear over the headphones
2. in the scenario, you are listening for approaching cars to ascertain if it safe to cross a somewhat busy street
3. you can expect to hear a single car approach you from either the left or the right at any given time, but never more than one approaching car will be present at a time;
4. these cars are moving at approximately 20 kmh, which could cause injury if you were to be struck
5. there is no stop light and there is only one street in front of you, it is not an intersection.
6. you are not really in danger!

Any questions so far?

Some of the approaching cars you will hear will sound like normal cars you might hear while walking in your neighborhood, or perhaps on a busy street in your city. Some of the cars may not sound like a normal, but quiet car, that has sound emitting from a loudspeaker mounted on it. In fact, most of the cars you will hear contain synthetic sounds designed to help you hear the car.

Any questions about this?

At this point I would like to play you samples of these approaching cars. The samples I am about to play are the same approaching cars you will hear during the experiment. However, in these demonstrations, each recording is of a car that will approach from a single direction. So during the demonstration, you will hear an example of each car you will hear in the experiment, but here, you will not hear a given car approach from the left and also from the right in a later example. Where as, in the experiment you will hear these cars approach from both directions. That is, all cars were recorded approaching from both directions at different times, but you won't hear all of these now. We are limiting the direction here in the interest of saving time. I would like to play these demonstrations for you and make sure you understand what you are listening for during the experiment.

Are you ready to listen to some examples?
Ready for the first example?

Open the folder 'demonstration sounds' and click on the first file in the list. Watch the playbar and when the sample terminates, ask them if it clearly sounded like an approaching car.

Was it clear to you that you heard an approaching car?
What direction did you hear it approach from?

They should answer this correctly, if not, repeat the sample after ensuring that the headphones are not reversed!
Do this for each demonstration sound. Only allow for one mistake, if they cannot clearly hear the direction of these sounds, they cannot hear or cannot understand what we are asking. Either way, they will likely provide bad data.
Once they have successfully heard all examples…

Background sound instructions:

So far, you have only heard the cars when there is no other background sound. It is clearly more dangerous for pedestrians when the surrounding environment is noisy, as it would be if a freeway were nearby. That is why we made the actual recordings next to a busy freeway. So, during the experiment, you will hear the same cars as you heard during the demonstration, but you will also hear a very busy freeway in the background. Furthermore, we chose to record on a rainy day to provide even more noise which perhaps can make for a very dangerous situation for a pedestrian who is reliant on sound.

What I would like to do now is play you a very short sample of the background sound so you know what you will be hearing during the training and the actual experiment.

Play the background demo file.

It should be clear now what you will be listening for in the actual experiment. However, I need you to remember that you just heard 11 different cars. You may have noticed that the presentations progressed so that a subsequent sample approached from the opposite direction of the previous example. This type of structure will not be maintained in the actual experiment. In the actual experiment, you could hear any of the 11 cars approach from either direction at any given time. Although, please remember that you will only hear one approaching car at a time. Also, after hearing the total passage of one approaching car, you may immediately hear another car approaching, or it may be up to 20 seconds before you hear another one. This is done so that the situation is more realistic, i.e. somewhat unpredictable.

Do you understand so far?
I will now play you a short example of the cars approaching according to the time structure that will be utilized in the training and actual experiment. We will do this just to ensure that you are not surprized or confused when you are hearing it in the critical part of the experiment.

Play them the random_practice file.

This is the type of timing and directional presentations you will hear in the real experiment, although it should be noted that the examples you just heard would only be a portion of the experiment. Furthermore, the timing, direction, and car will continuously vary throughout the experiment. Still, please keep in mind that you will never hear more than one approaching car at a time, and that the cars will only approach from the left or the right.

Any questions?

Task instructions
I would now like to tell you more about your task in the experiment.
During the experiment, when you are sure that you hear an approaching car, you respond by pressing either the spacebar (show them the space bar) with your left hand to indicate you are sure you hear the car approaching from the left. Likewise, when you hear a car approaching from the right, tell me by pressing the enter button with your right hand.

(Have them place their fingers on the buttons as they will during the experiment)

If you hear a car approaching from your left, which button will you press?
From your right?
Good.

It is very important that you try to imagine that you are actually trying to cross the street, and that if you make a mistake, you could be injured. Of course, there is no real danger here, but it will not help us if our participants only see their task as pushing a button when they hear something. This is why we emphasize that you should only respond as quickly as possible when you think you hear a car. However, it is also important that you try to hear from which direction the car is coming from as previously described. In short, we need you to respond as quickly and accurately as possible. Just keep in mind that it is more important to first detect the sound, as it would be in a real-life situation. It is okay to make mistakes, just do your best.

Any questions about that?

We are just about ready to try a short training session, but I need to give you just a little more information.

Training instructions

Are you ready to begin the training session?

Before you begin, I would like to tell you that during the training session and the actual experiment, I will leave the room so you will be free of distraction. The experiment will proceed in blocks. After each block, I will come back in to see how you are doing. Ready for the training block?

Open the 'practice' folder and click on the experiment program when they are ready.
Monitor their progress on the screen to ensure they understand what they are doing.

Once it has concluded…

Great! It looks like you are ready for the actual experiment!
Do you have any questions before we begin?
Remember that after each block, you will be given a short break in case you need a drink etc.
So, if you feel ready, let's get started!

Block 1 instructions
Open 'block 1' folder and click the program file when they are ready.
Monitor their progress.

When the first block concludes…

How are you doing?

You now have a few minutes to take a break if you like. Feel free to get a drink, use the rest room etc.

Any questions before completing the final part of the experiment?

Block 2 instructions:
Open 'block 2' folder and click the program file when they are ready.
Monitor their progress.

When the second block concludes…

Great job! You have successfully finished the experiment. Now that you are finished, I would like to tell you more about our project.

Debriefing:

You may have noticed that some cars were more difficult to hear than others. These were likely electric vehicles. With the decrease in noise produced by hybrid and electric vehicles, urban environments are increasingly getting quieter as well. In fact, it certainly is an overall benefit to have quiet vehicles in terms of the reduction of noise pollution. However, it is also the case that pedestrians and even animals are at risk as a result. This is especially true if the drivers are not paying close attention to the road. With the ever increasing technology of mobile phones and computer interfaces within vehicles, drivers are markedly more distracted than they have ever been. The goal of this project is to try to find the important aspects of synthetic sounds that may be added to quiet cars that are at least as effective as the sound of a normal car (ICE). Furthermore, we hope that there is a way to use a much quieter sound while remaining just as effective as a normal car. These sounds can be emitted through an intelligent sound distribution system for optimal performance. Thanks to yours and other participants help, we may be able to help keep our environment quiet and safe.

Do you have any questions?

If you have any questions regarding your performance or your experiment in general, I would be happy to provide details once we have analyzed your data. Please feel free to contact me via email. Thank you again for helping us, your participation is appreciated.