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.		
Deliverable Title	Parameter Selection and Stimuli Design Proposal	
Dissemination level	Public/Confidential/Restricted	
Written By	Ryan Robart (INSA)	Date
Checked by	Etienne Parizet (INSA)	Date
Approved by	Name techn. Coord. (Organisation)	Date
Issue date	Date	

Executive summary

One of the primary challenges of this project is the apparent lack of research that has focused on dynamic sound source localization, segregation within dynamic sonic environments. The lack of research is likely due to the overwhelming choice of variables and logistical concerns regarding control variables. Despite the lack of research, there is little doubt that there is need for this type of work and that the academic and industrial communities will benefit from carefully design studies that are based on empirical research. When researchers are faced with a lack of central research, they often look to relevant empirical literature that is associated with the central topic in order to derive converging evidence that can guide them in their research design. Fortunately, there are numerous lines of research that may be instructive for our stimuli design. In this deliverable, such an approach is taken by reviewing various domains of empirical literature that we propose should be the basis for our stimuli design. The state of the art is considered throughout, and certain sonic parameters are proposed to be critical for our design(s). The parameter selection is then outlined and described in detail. Taguchi's (1987) Orthogonal Table methods for stimuli selection are reviewed and applied to our proposed design parameters. Finally, the development and utility of a sound design tool (eVADER synth) is briefly discussed.

Contents

1 Introduction (Literature Review)

1.1 State of the Art

While synthesized automobile sounds are fairly common in video game design and perceptual research utilizing driving simulations, there is a surprising lack research and development for the practical use of such sounds for quiet vehicles. The accident data and the increasing number or electric and hybrid vehicles have provided some background to begin developing candidate sounds (Tabata, Konet, Kanuma, 2011; Chamard & Rousarie 2012; Misdarlis, Levallois, Locqueteau, 2012). In fact, Japan has provided a reasonable protocol for the design of such sounds (see Tabata et al, 2011). There are 3 concerns that are the basis for their protocol:

- Sounds must be detectable to pedestrians
- Sounds must be quiet in the car
- Sounds must not add annoying noise to environment

Clearly, a paradoxical relationship exists between 1 and 3. Furthermore, these guidelines do not provide any aid in solving the paradox. Nonetheless, Nissan has progressively pursued the challenge and have provided the most useful protocol for the design of such sounds (Tabata et al, 2011).

- Sound is recognized as a vehicle
- Pitch is proportional to the speed of the vehicle
- Sound level should be similar to an internal combustion engine (ICE)
- Sound has a futuristic image
- Sound should be audible even to elderly people with hearing damage, without adding noise to environment
- Sound system can be temporarily disabled.

Again, there is an obvious paradox in requirement 5. Still, Nissan has developed and use a synthetic sound in some of their EVs using this protocol (Tabata et al, 2011). Though, it seems the sound is safe for pedestrians, it is apparently not marketable as an ascetic replacement.

It is commendable that Nissan has recognized the importance of scholarship in the research, design and implementation of their replacement sound (Tabata et al, 2011). For example, their 'twin peaks one dip' spectral envelope was chosen based on urban noise research (peaks at 1 kHz) as well as human sensitivity research. Furthermore, Tabata et al (2011) notes that spectro-temporal modulation is likely an important sound feature for listener segregation of a sound stream from a background of noise as well as for vehicle classification and the recovery of dynamic properties of vehicle motion. Although, it should be noted that little literature (if any) exists on that specific topic. In fact, there is a surprisingly lack of references to the relevant cognitive literature on attention, perception and memory. Similarly, it seems clear that masking is a primary concern, which is absent from most of the work considered to be the state of the art. Still, Nissan (Tabata et al, 2011) has clearly made important steps in deriving an empirically based sound design protocol, it seems clear that more research is needed to complete such a protocol.

Similarly, a recent report on more complex synthetic sounds reflects great ambition and even greater abandon for scholarship (Chambard et al, 2012). Like Tabata (2011), Chambard (2012) developed candidate spectral ranges and spectro-temporal modulation for synthetic

vehicle sounds. However, there are no citations to any relevant work that could be the basis for their stimuli design. Furthermore, the experimental methodology is unclear and the results are lacking in description. Therefore, the study does little in progressing science or explaining the state of the art.

Other recent work done by Misdariis et al (2012) does seem to be based on the empirical literature. Though, in the report the citations of seemingly very relevant work (e.g. Kerber et al, 2008) tare not clearly described. Moreover, most citations are unavailable online which can only make such references questionable. Typically though, the estute reader may derive the important aspects of referenced work based on the pragmatic report of the methods supplied by the author. Frustratingly, there does not seem to be any empirical basis for Misdariis et al (2012) parameter selection. A reference to futuristic car sounds from popular science fiction films, suggests, to this author, that these sounds are primarily designed for ascetic concerns. However, based on the data, it appears that their stimuli did produce fairly good detectability (Misdariis et al, 2012). Still, the general methodology is unclear, as are the data they used for their statistics. For example, it is impossible to interpret the results of an ANOVA without seeing the degrees of freedom and the F-score (see Rosenthal & Rosnow, 2008). A further concern stems from a reference to the Shafer (1993) acoustic ecology guidelines, which apparently was the basis for their sound design protocol (Misdariis et al 2012). Once again, the literature cited is not described, even though it seems like relevant listing guidelines would be fairly intuitive. However, Misdariis et al (2012) do list *their* guidelines that are supposedly based on the Shafer (1993).

The Misdariis et al protocol (2010) that was based on Schafer (1993) were as follows:

- Sounds in the low frequency range (20-100 Hz) should be avoided.
- Sounds should have some energy around 1000 Hz
- The sound should be a clean, static, ordered sound that can emerge from the soundscape
- The sound should be detectable:
- Should contain a dense and rich layer of sound at higher frequencies (around 3000 Hz.)
- The richness should consist of swooshes of filtered / granulated white noise) in order to be differentiated from other traditional components emitting in the same frequency zone (e.g. squealing noise of trucks' brakes).
- Sound should have a low intensity level

It could be that the features in this list are essential for any sound to satisfy the outlined requirements for a suitable alert sound in a modified vehicle. From an empirical perspective, it certainly is unclear why the features in the Misdariis et al protocol (2012) are even important. This is because there is essentially no reference to research that validates these claims. To be fair, Misdariis et al (2012) does cite some perceptual research, but it is difficult to understand where the motivation for their stimulus design originates from a cognitive perspective. Of course, their work could be extremely valuable, but without a pragmatic process of research design, it is unclear what the results really tell us. Certainly, it further illuminates the need for a link to be formed between the cognitive literature and the sound design choices made by car manufactures and researchers alike.

It is clear that there has been some effort in determining some form of protocol for this type of sound design (Tabata et al, 2011; Chamard et al, 2012; Misdariis et al 2012). While the examples reviewed above may seem instructive for our purposes, it is arguable that they simply beg the question(s) regarding empirical results. It is imperative to have clear foundations for research and stimulus design. The only possibility for such a basis is Shafer (1993), which frankly is not a scientific theory. It is simply not enough to reference the sensitivity literature alone, as we have little evidence to support any conclusions about human sensitivity in an urban environment. Therefore, it is necessary to look at both the sensitivity literature as well as the cognitive literature to support our choices for design. The next section of this literature review will summarize a theoretical framework that can be the basis for our empirically based protocol for stimuli design proposal.

1.2 Auditory Scene Analysis

Auditory Scene Analysis (Bregman 1990), the most comprehensive and well-known theory of auditory perception, was developed to in the interest of applying seemingly intangible psychophysical studies to common perceptual phenomenology (Bregman 2001, Neuhoff et al, 2002). Explaining the ability to 'hear out' certain collections of sounds from a cacophony of sonic events has been one of the primary goals of ASA (Bregman, 1990; 2001, see also Carolyn, 2004). From the perspective of ASA, the ability of a listener to perceive certain collections of sounds as groups is due to a process of perceptual organization known as stream segregation and/or grouping (Bregman, 1990, see Moore 2003 for a succinct review). Often, the results of psychophysical studies are argued to have very little ecological validity because of the unnatural presentation of unnatural sounds (Bregman, 1990). A sine wave tone, for example, does not occur in the natural world unless a human constructs a machine that produces one. The guidelines outlined in the above section demand that candidate synthetic sounds must be recognizable as a car (Tabata, 2012; Chamard et al, 2012; Misdariis et al, 2012). ASA can provide a link between simple sounds and what could be essential for the classification of complex sound structures like that of an automobile engine (Bregman, 1990). This is particularly interesting for our purposes, as our goal is to practically combine empirically derived results from very controlled experiments that have often used simple sounds. Next, the relevant constructs of ASA associated with more basic processes are reviewed

1.2.1 Grouping and Segregation

Auditory grouping/stream segregation is the measured by the number of reported sound sources a listener hears while listening to many, competing sounds (see Bregman, 1990). The research conducted on auditory grouping has attempted to parametrically define conditions that facilitate and/or diminish such grouping(s) (Bregman, 1990; and see Best,Gallun, Carlile & Shinn-Cunningham, 2006). The critical points when sounds are segregated into separate groups are referred to as the temporal coherence or fission boundary (Bregman, 1990 p. 60). This is crucial for both sequential (sounds emitted in a sequence) and simultaneous (sounds are simultaneously emitted) types of segregation/grouping.

According to ASA (Bregman, 1990), the temporal coherence and fission boundaries are determined by automatic, i.e. primitive, cognitive processes. In other words, the listener cannot help but segregate sounds into two groups –or streams– if the sounds are sufficiently dissimilar in frequency (sequential and simultaneous types)(Bregman, 1990). However, it is has been established that even subtle but uniform frequency and amplitude modulation reduces the segregation (see Moore 2003; Bregman, 1990 chapter 4). The 'state of the art' literature persistently mentions terms associated with figure/ground segregation (Tabata et al, 2011; Chamard et al 2012; Misdariis et al, 2012), but nearly fails to reference any theoretical basis to explain the process. The construct of stream segregation is pervasive in the whole of the auditory perception research, and is a strong argument for any effects discussed in the state of the art literature reviewed above (Tabata et al, 2011; Chamard et al 2012; Misdariis et al, 2012; and see Carolyn, 2004). It remains an open question though, what specific features of any candidate sound could be optimal for our stimuli requirements.

1.2.2 Sequential Grouping/Segregation

Sequential grouping or 'musical grouping' has been the focus of a wealth of literature over the last 50 years (Bregman, 1990). These types of sounds are really just sequences of sounds presented at a rapid rate. The grouping/segregation occurs when the listener stops hearing the sounds as what they are (many), and allocates certain sounds to groups in their mind. As noted above, frequency similarity is considered a primary source for grouping, while frequency dissimilarity is a primary source for segregation. Though, temporal aspects of the presentation of tones are also primary. For example, sounds presented at either extremely rapid (e.g. 20 per second) or extremely slow rates (1 per second) aren't usually segregated from dissimilar, sequentially presented sounds (Bregman, 1990). The durations of the inter-stimulus intervals (ISIs), or the duration of silence between each tone in a sequence are also critical. Research has generally shown that irregular ISIs that are either too short (e.g. 5-10 ms) or too long (500 ms) can play a large role in stream segregation (see Bregman, 1990, Chapter 2). In other words, if there are intermittent irregularities in the timing of the sounds in a sequence, listeners will over whelming segregate the sounds into separate streams. On the other hand, if a tone sequence contains tones that are sufficiently similar and/or have a suitable duration(s) (e.g. 50-200ms) along with regular ISIs, a listener will probably hear the sequence as one coherent stream. Concurrently, the release from masking literature (discussed in more detail later; Moore, 2003), suggests that onset asynchronies as small as 5 ms markedly reduces the threshold for detection of tones when presented with a much louder masker tone (see Moore 2003 for a review). This demonstrates the importance of the temporal aspects of sequential presentations as opposed to simultaneous presentations of sounds. These results argue for a sequential-based sound structure in our list of considerations for stimuli design.

Returning to the 'state of the art' literature, the sound used by Nissan could be argued to contain modulation parameters that give their sound a sequence-like structure (Tabata et al, 2011; Bregman, 1990). That is, the amplitude modulation creates systematic changes in the level of different frequencies such that it could be characterized as a sequence. It should be mentioned though, that those modulations may be too fast based on the requirements of stream formation as outlined by Bregman (1990) see fig below.

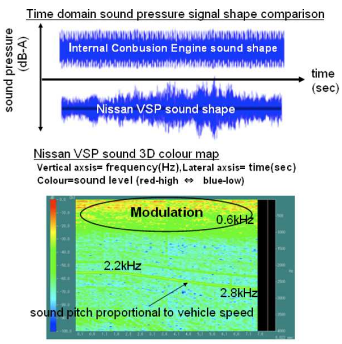


Figure 1-1. Tabata et al.'s (2011) figure demonstrating frequency content of an ICE, and their synthesized sound.

The modulation occurs at 600 times per second according to the bottom part of the figure. However, the top portion of the figure indicates that there is some amplitude modulation that is much slower. If the waveforms displayed on top occur over 8 sec; clearly there is some slow amplitude modulation that could be considered as a sequential structure (Bregman, 1990).

Misdariis et al (2012) also provides sonograms that seem to illuminate a sequential (spectro-temporal modulation) structure see figure below.

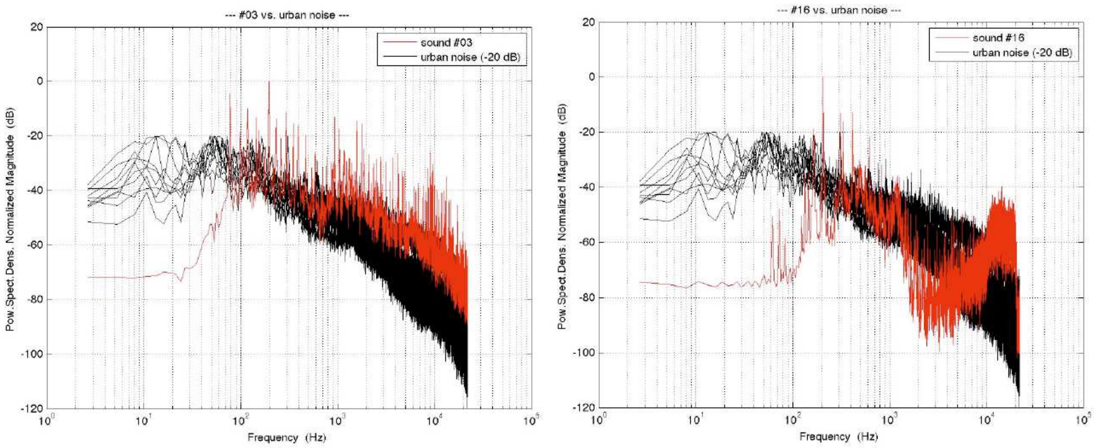


Figure 1-2 Spectral emergence figure taken from Misdariis et al (2012).

These graphs seem to convey that spectral emergence is due to some kind of modulation, but it is not clear how these sounds unfold over time. This is particularly important for pedagogical reasons because the time domain could provide some information that may be focus of previous studies. Since we do not know how or why these sounds were made, the spectro-temporal dynamics, or even what they sound like, this research does little to progress the state of the art. Still, at least we might assume that there is some use of a dynamic, spectro-temporal sequence-like structure, which we know from ASA to be a highly useful parametric choice for such sounds.

It appears that sounds used by Chamard et al (2012) could be lacking in this respect. Based on the sonograms provided by Chamard et al (2012), the sound resembles filtered noise with a very slow amplitude modulation (see fig below).

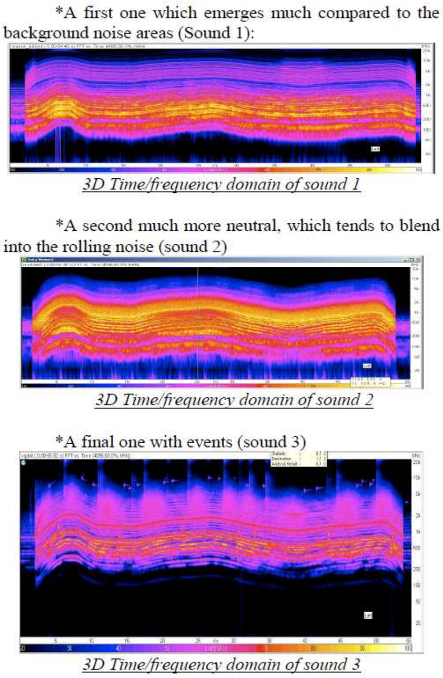


Figure 1-3 Sonograms taken from Chamard et al (2012) comparing the spectro-temporal content of 3 of their synthesized sounds.

There may be evidence of slow amplitude modulation based on periodic gaps in the sonograms, but based on the course used, it is difficult to interpret the sonograms in terms of what the vehicle is doing at any given time. Hence, we can't be sure how much of the variation shown in the sonogram is due to the modulation with in the sound design or due to the dynamics of the vehicle. Generally though, the coverage of the frequency range in the smear indicates a fairly temporally uniform complex sound with some variation in amplitude as indicated by color changes. This is more akin to simultaneously presented sounds (Bregman 1990), and is likely not as useful as tool for facilitating segregation between many sound sources (see Moore, 2003). This is likely because a sequential structure offers much more opportunity for manipulation and therefore contains global manipulations that will be used by listeners for grouping/segregation (see Bregman, 1990 chapter 6 'old + new heuristic'). Yet, it is still unclear if such sounds are an ecologically valid choice, as we have only discussed primitive processes in depth. The next section will review the literature associated with grouping/segregation based on schemas, a higher-order cognitive representation.

1.2.3 Schema-based Grouping/Segregation

Schema based segregation is proposed to be based on experience as opposed to primitive mechanisms of mental processes (like reflexes) (Bregman 1990). Schema-based mental processing is very difficult to study, as it involves many processes such as memory, attention, perception, problem solving etc. As a result, the literature regarding the schema

construct is far more limited than the literature for primitive construct. It is most disturbing because schema based segregation must occupy the majority of everyday listening (Bregman, 1990). Still, the assumptions of schema-based segregation remain very useful for our purposes here. For example, the visually impaired have been found to be insistent that modified vehicles should sound like real vehicles (Robart & Rosenblum 2009). This assertion is in line with the assumptions of schema-based segregation, as it points to an existing schema for car-like sounds (Bregman, 1990). Furthermore, research has shown that performance regarding learning and memory and spatial segregation is improved when familiar sounds are used as stimuli (see Edworthy & Hards 1997 for examples of this). Based on these results and the demands of the guidelines reviewed above, it seems that the most informative and easily learned stimulus will bear some resemblance to engine sounds. Still, it is unclear what is required for schematic classification. This idea will be discussed in more detail in a later section titled 'Lessons from the Cognitive Literature'.

Returning to primitive segregation (sequential grouping), it is noteworthy that some research has showed that grouping of simple tones varies according spatial location, especially when the tones are moving (Robart, 2010). This suggests that differences in spatial location and dynamic properties of motion such as trajectory can be enough to negate grouping by frequency. It is well known in the sensitivity research that interaural time differences (ITD), interaural level differences (ILD), and the natural phase effects of wave propagation can help explain these results (see Moore, 2003; Bregman, 1990). This is particularly instructive for our purposes here because it shows that listener confusion regarding the grouping of similar sounds, such as 2 or more modified electric vehicles with similar alert sounds, will be negated if the cars are sufficiently separated in space. More importantly, the dynamic aspects of motion, such as trajectory and speed, may be sufficient for listeners to segregate the sound sources. Still, this is suggestive that if these simple sounds could be made to sound more car-like, it could improve segregation and localization. It is not understood if simultaneous sound segregation offers the same benefits. It is an empirical question, but it seems unlikely that extensive spectral complexity in a simultaneous sound presentation (with little sequential structure) can provide enough variation for segregation from competing sound sources like other electric vehicles (see Moore, 2003). Certainly, an extensively complex simultaneously presented sound such as filtered noise (Chamard et al, 2012) may be enough for good figure/ground segregation when presented alone (Bregman, 1990). It seems likely that much less noise and much more and much slower amplitude modulation would likely lead to better segregation when presented with competing sound sources, such as other electric vehicles with synthetic sounds. A central concern regarding noise based simultaneous sounds is masking, which is the subject of the next section.

While this summary of ASA is admittedly very brief, the most pertinent constructs were summarized. It is essential to understand the differences between primitive and schematic processes and the empirical evidence associated with them. It is likely that the whole of our perceptual experience is based on schematic processes. This is why it is problematic to apply evidence regarding primitive processes (like frequency sensitivity) to support decisions for practical applications. ASA provides a theoretical framework by which we can combine such evidence; along with evidence associated with schematic process into converging evidence to support our stimuli design protocol. The next section will concentrate on detection and sensation with the goal of extracting meaningful results that are instructive as to our endeavor.

1.3 Human Sensitivity and Masking

1.3.1 Spectral Complexity

There is a vast literature on human sensitivity and localization of various sounds such as sinusoidal tones and filtered noise (Zahorik et al 2004; Middlebrooks and Green 1991; Hafter and Trahiotis 1997). Furthermore, long duration (>200ms) sounds are often chosen for motion detection experiments (see Grantham et al 2003 for a review). However, it is difficult to interpret results from these types of controlled studies in an environmental context because the sensitivity data often is only applicable to a given set of circumstances, such as an anechoic environment. Still, this literature can offer some potentially important constraints. For example, research on the detection of motion and source segregation has most often made use of complex sounds (e.g. filtered noise) (see Zahorik et. al. 2004; Grantham et.al. 2003). Although there have been other studies that have shown comparable performance using simple tones (Grantham et al 1986), it is likely that the choice to use complex stimuli is based on physics (Zahorik et al 2004). It has been established that human physiology can define frequency sensitivity and selectivity (Moore 2003). More specifically, the shape of our ears, head and shoulders act as a set of filters (head related transfer functions (HRTFs). HRTFs can be potentially confounding when the location of a sine wave sound is varied (see Grantham 2003 and Zahorik et al 2004). This is due to sound wave propagation around the body. By using complex stimuli like filtered noise, the potential for these types of problems are minimized because of the numerous frequencies that exist within the noise. This can be taken as an argument for complexity of spectral content in our list of stimulus design considerations.

While complexity and noise are not mutually exclusive, it is possible to create complexity amongst within a simple tone complex by way of modulation (Moore, 2003; Bregman, 1990). Considering the sonic complexity of a typical urban environment, it seems prudent to focus on the literature involving localization or detection of sounds that are embedded in other sounds (e.g. harmonic complexes or noise) (Moore, 2003). Certainly, the rich literature on release from masking can provide some lessons for a list of considerations our stimulus design.

1.3.2 Masking

It is widely known that the threshold for detection of certain sounds (e.g. 1khz sine wave, narrow band noise, etc.) can be depend on the spl and bandwidth of noise that it is presented with (1 kHz centered narrow band noise) (see Moore 2003 for a review). This has been concluded to be the result of the existence of critical bands, which are thought to be due to frequency selective neurons. As reviewed by Moore (2003), the exact number of and size of the critical bands (filters) is generally uncertain. There has been a plethora of research producing masking effects with the goal of determining the shape(s) of the filters, and the general conclusion is that the we can only assume that it is non linear. Regardless, we can assume that masking is a royal concern for our purposes, as we have the unique challenge (and privilege) of designing stimuli that should be as quiet as possible while reliably avoiding masking from louder sound sources, as well as a noisy background.

As previously discussed, the sounds used by Chamard et al (2012) are certainly wide band noise stimuli. It is also probably the case that masking will occur if other vehicles have a similar spectro-temporal structure (Moore, 2003). More specifically, wide band uniform structured sounds are probably poor candidates for sounds that require source segregation. This is an empirical question, but the solution seems to be in the literature (Moore, 2003). When considering the figures provided by Nissan (Tabata et al, 2011), the modulation is bit more obvious in the rectangular gaps in the sonogram. It could be that the shifts in pitch that

vary according to speed, can probably help with masking release (Moore, 2003) Still, it is likely the case that a stationary Nissan would mask an approaching modified vehicle with the same sound. It seems that there must be some kind of modulation, even when the car is stationary. The co-modulation release from masking literature could provide some empirical evidence that may help us strengthen the foundations of our stimuli design.

1.3.3 Co-modulated Release from Masking

Interestingly, modulating the amplitude of the background decreases the threshold for detection of a sound (see Moore, 2003). This has been shown with multiple bandwidths when an additional flanking sound (outside of the critical band) is added and modulated with the masking band. Furthermore, this effect is robust when the flanking sound is presented to the opposite ear. This has implications for top-down processes (schema-based) overcoming bottom-up processes (primitive). These compelling results flirt with the possibility of illuminating the boundary between sensation and perception/attention. For our purposes here, it can be considered an argument for amplitude co-modulation as a source of complexity. However, it should be noted that the task of the listeners in these experiments was to detect the *unmodulated source* (Moore, 2003). Still, it seems clear that the segregation occurs as a result of the modulation. In other words, the listener might have compared the uniform change in amplitude structure to the unchanging amplitude structure in order to complete the task (see Bregman, 1990 chapter 6 'old+new heuristic'). It is prudent to note that this has not *explicitly* been shown empirically. However, the state of the art literature (Tabata et al, 2011; Chamard et al, 2012; Misdariis et al, 2012) clearly make reference to similar evidence, it is not clear based on their citations, where such evidence exists. While it speaks to the progressive spirit of those who have bravely pursued this work, it could be argued that we simply do not know a real state of the art. Plainly, it may be our task to clearly define the state of the art by conducting carefully designed experiments that are pedagogically and epistemologically based. Pedagogy requires that we reference the epistemological cognitive literature to help in this endeavor.

1.4 Lessons from the Cognitive Literature

1.4.1 Salience

Turning to the cognitive literature, studies of attention can certainly be instructional for our considerations of stimuli design. While somewhat controversial, it is generally accepted that bottom-up (primitive) attention driven by feature selection and integration (see Treisman and Gelade 1980). An extremely important concept in the attention literature is the construct of salience (Kaysar et al 2005). Salience is a tricky construct as features, such as frequency, can have salient features, such as intensity. Another elusive aspect of salience is that it can drive primitive processes as well as top-down (schema-based) processes (Kaysar et al 2005). That is, schema-based segregation as outlined by ASA holds that certain classifications of sounds (schemas) have collections of features (spectro-temporal relationships) that are common to most sounds within the class (Bregman, 1990). As previously discussed, ASA would contend that vehicles such as cars have common spectro-temporal features which are likely to be salient for their segregation from a background as well as other sound sources (Bregman 1990). Fortunately, the candidate sound parameters discussed so far may all be allocated to a vehicle sound class (see Poirson et al 2010; Wang 2008). However, since we are not trying to re-create engine sounds per se, it is imperative to ensure that our spectral selection contain features that are known to be salient at the bottom-up level.

1.4.2 Harmonicity and the Salience of Enharmonic Variations

A group of studies utilizing complex yet noise-free stimuli has produced results that may be useful in our endeavor. More specifically the literature shows that attention is modulated by the degree to which a sound sequence or a simultaneously presented tone complex is harmonic (McDonald & Alain 2005; Moore, 2003) . For simultaneous presentations, the results generally show that mistuning a single harmonic among 3-10 can facilitate robust segregation (Sussman et al, 2005; McDonald & Alain, 2005; Moore, 2003; Bregman, 1990). This effect was amplified as a function of location and intensity (McDonald & Alain 2005; Moore, 2003; also see Woods et al, 2001). Both are important factors to consider for an environment that contains multiple versions of the same sound that vary according to spatial location, trajectory and acceleration (Gaver, 1993). Taken together with the assumptions of ASA, and the knowledge of the complexity of our sonic world, these conclusions can be formulated into an argument for a *mostly* harmonic spectral structure in our list of stimuli design considerations. However, there may be more to find in work that has made use of dissonance within sound/sound sequence structure.

There has been some interesting work focused on attention and auditory streaming by measuring event related potentials (ERPs) associated with dissonance (McDonald & Alain, 2005; Sussman et al, 2005). The most common ERP in attention research is known as the mismatch negativity response (MMN). Many studies have proven that the MMN response occurs only when the brain detects a change in the stimulus array. The MMN is elicited even if the observer is unaware of any changes in the stimulus array. Therefore, the MMN is argued to be an indicator of a pre-attentive process, or an early attentive process. This is important for our purposes, because it is often the case that pedestrians are not paying attention when walking. It is known that stimuli that elicit a MMN response lead to faster reaction times when the observer becomes aware of changes in stimuli. The most notable work conducted by McDonald & Alain (2005) and Sussman et al (2005) focused on ERPs to analyze the level of attention (preattentive/executive) dedicated to the sequential segregation based on mistuning of tones in a sequence as opposed to a simultaneous presentation. Interestingly, the Sussman et al (2005) experiments used manipulations of repetition within the sequence structure to produce mistuning. For example, imagine the global sequence structure was as follows:

Low (1), Medium (1), High (1), Low (2), Medium (2), High (2), Low (3), Medium (2), High (3)…

This structure included random repetitions of certain sounds at a predetermined percentage see figure pictured on the next page.

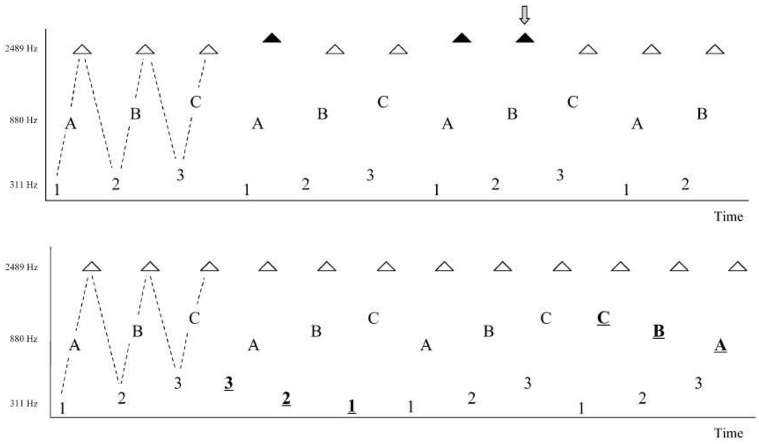


Figure 1. Stimulus paradigm for Experiment 1: schematic diagram showing the presentation of the alternating high- (top row), middle- (middle row), and low- (bottom row) frequency range tones. The top panel shows the frequent high tones (white triangles) and the infrequent higher pitched tones (black triangles), along with the standard three-tone rising patterns occurring in the middle range (denoted with different letters) and in the low one (denoted with different numbers). The target for the attention-based task is designated with an arrow (two successive higher pitched tones). The bottom panel shows the three-tone pattern violations (reversal of order) occurring within the middle- and low-frequency range tones (denoted in boldface and underlined).

Figure 1-4 Illustration of sequential stimuli taken from Sussman et al (2005).

The most relevant result was that MMN was elicited when there were variations to the global structure when the listeners were not attending to sounds at all (Sussman et al 2005). This was accomplished by presenting the stimuli while giving the listeners a visual task (matched for difficulty). It was concluded that these results were due to a pre-attentive segregation process. Along with the simultaneous segregation results (see Bregman, 1990; Moore, 2003), these results present a compelling argument for the salience of slight enharmonic variations in our list of stimuli design considerations.

So far, the previously highlighted features we are proposing can all potentially be found in the sound of an average engine (see Poirson et al 2010) . Of course, the ideal engine would be a completely harmonic machine. Yet, our experience can confirm that very few vehicles are free from imperfections manifested by slight enharmonic sound. The attention literature suggests that pronounced variations of harmonicity based on vehicle dynamics is a good solution for detection and segregation (Sussman et al, 2005; McDonald & Alain 2005). A counter argument might be, that this will only be true for moving vehicles. That is, it may be very difficult for a visually impaired person to discern between multiple stationary modified vehicles (at an intersection), if there is no variation between them. Therefore, it seems that it may make sense for each make and model have some slight enharmonic variation in the spectrum, even while stationary. Along with temporal dissimilarity, these harmonic parameters seem like a salient source for listener segregation and could be source for constant modulation to increase the informative nature of the sound. However, it is prudent to review the literature regarding alarm sounds, as they are certainly included in our environmental schemata.

1.4.3 Alarm Sound Schemata

As prescribed by ASA, it is likely that the average listener has schemata for common sounds (Bregman, 1990). Alarm sounds are omnipresent in an urban environment, and have been the subject of numerous studies in the cognitive literature (see Parizet 2011 for a review). Most research has focused on reaction time (detection), learning and memory. However, there are at least 3 limitations for our application of this literature. The first is that most alarm sounds are designed to be loud and annoying. Presumably, this is done in order to facilitate a reaction. Often, an alarm is only temporary. Such alarms will either terminate at regular intervals, like that of a crosswalk, or will terminate as the result of a reaction such as seat belt alarms. The second limitation is that continuous alarm sounds are generally found in areas that require constant monitoring like cockpits or hospital rooms; not on roadway. This points to the novelty of the goal of our project. A third limitation is that most existing alarms, hence the focus stimuli of alarm sound research, has relied on unnatural sounds

1.4.4 A Realistic Advantage

Interestingly, there have been studies comparing abstract, real, and synthesized sounds on their differential effects measured by memory and reaction time (Edworthy & Hards, 1997; Stanton & Edworthy 1998; Sued et al, 2010). It has generally been found that natural sounds are easier to learn and remember and produce equitable reaction times to abstract sounds (see Parizet, 2006 for a review). This is at least partly in line with the schema-based assumptions of ASA (Bregman, 1990). Though, it has been argued that natural sounds are simply more acoustically complex which may point to featural salience as opposed to schematic contributions to performance (Sueid et al, 2010). Regardless, the results of such experiments seem to support our proposed parameters, and the requirements for our design (Tabata et al, 2011). Furthermore, it is well known that measures of urgency are highly correlated with temporal structure (see Parizet, 2006 for a review). Even more compelling, is that it is clear that urgency has been measured to increase as the temporal structure increases (e.g. ISI or frequency of modulation etc.) (Patterson, 1985). Certainly, it could be argued that our experience with everyday sounds, such as motors, could account for such robust results (Bregman, 1990). That is, it is likely that we have schemata regarding the acceleration of motion as conveyed by an increase in frequency as well as a decrease in time between repetitive sounds such as impulsive sounds.

1.4.5 The Patterson & Mayfield Prototype

Patterson & Mayfield (1990) provided a prototypical ergonomic sound that was meant for use in a train or an airplane. While it was not designed specifically for the demands of our endeavor, it is strikingly demonstrative of a stimulus designed according to our proposed parameters (see figure on the following page).

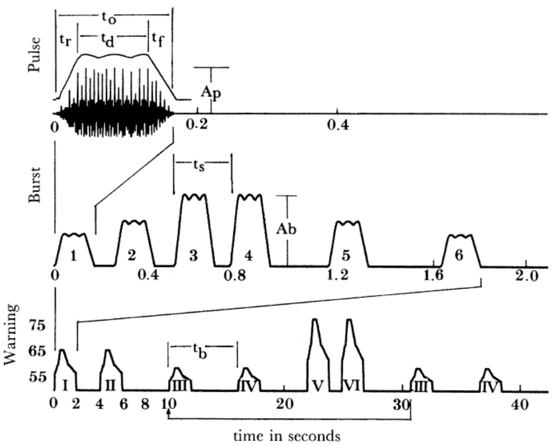


FIGURE 3. The modules of a prototype warning sound: the sound pulse is an acoustic wave with rounded onsets and offsets and a distinctive spectrum; the burst is a set of pulses with a distinctive rhythm and pitch contour; the complete warning sound is a set of bursts with varying intensity and urgency.

Figure 1-5 Illustration of a prototype alarm sound taken from Patterson & Mayfield (1990).

It seems that this prototype has all of the dimensions mentioned above. However, it is unclear if this particular example would sound like a car. Yet, by comparing the Patterson & Mayfield (1990) prototype waveform to waveforms provided by Tabata (2011) it could be argued that there is a visual similarity between the prototype and the Nissan waveform. The most compelling aspect of this prototype structure is that it has multiple dimensions that are candidates for manipulation. At the very least, it is a good reference to a possible result of a combination of our proposed parameters. Of course, this would only be the case if there were compression along the dimension of time. While it is true that realistic or natural sounds are clearly preferable for source identification and seemingly localization (see Parizet, 2011), we have seen that pulsing sounds and irregular ISIs can help with detection and derivation of urgency (Patterson 1985; Patterson & Mayfield, 1990). This suggests that some aspects of seemingly abstract sounds could be useful in our designs.

1.4.6 Are Abstract Sounds Underestimated?

Catchpole et al (2004) examined the localizability and identification of tonal and noise complexes in very informative group of studies. They tested performance under 3 relevant stimulus dimensions.

1. Spectral content:
    a. Tones
    b. Noise
    c. Combined
2. Phase limited Frequency modulation
    a. Up-sweep
    b. Down-sweep
    c. None
3. Spatial location
    a. 1
    b. 2
    c. 3

Also noteworthy, duration (250 ms) and amplitude (80-87 dB) were held constant for all stimuli. The large variations in dB are likely due to the frequency modulation. The results showed that filtered noise between 1-3 kHz produced localization that was not significantly different ($p > .05$) than broadband noise. Furthermore, they found that the notched noise combined with a tone complex with an upswept frequency modulation provided the best localization and identification. It is likely that the stimuli used in these experiments do not resemble the sound of a car. However, it is prudent to note that the stimuli are impulsive in nature and have a burst-like quality that is known to be binaurally informative. Though it is likely if these stimuli were to loop as they played, they could sound like a motor of some sort (Gaver, 1993). But it is interesting to imagine what performance would be for looping patterns. More importantly, the Catchpole et al (2004) results and the results from Patterson et al (1985;1990) could be the basis for an argument for periodic sharp onsets in our list of considerations for stimuli design (see Patterson & Mayfield 1992). That is, it seems that periodic bursts of energy are naturally informative for navigational purposes. Further more, it seems that such sounds will be salient for attention, especially if they are not constant.

2 Stimuli Design Proposal

2.1 Review of Research Design and Taguchi's Orthogonal Tables

The optimal solution for our project would be to develop a model that can accurately predict listener performance based on certain parameters. Typically, a linear regression model, or a structural equation model would be used for such an endeavor. Since we do not have a data set from which to draw from, we must gather our data. A factorial design is preferable for experiments with categorical variables (Rosenthal & Rosnow, 2008). However, continuous variables are more informative but often not available. Our goal is to attempt to find the parameters that might lead us to such variables, but the possibilities can be overwhelming. Furthermore, we have the unique challenge of using very long stimulus durations, which severely limits the number of trials we can include in our experiments. By using a repeated measures design, we increase our statistical power by reducing the likelihood of violations of assumptions associated with fewer trials and between subject effects. Still, we are limited by seemingly innumerable stimulus options, as the ANOVA structure is best utilized when only one variable (sound feature) is manipulated at a time. That is, the complexity of our sound stimuli will such that we likely could not appropriately account for potentially systematic variance partitions. Furthermore, we are uncertain of whether any effects would be random or fixed (Rosenthal & Rosnow, 2008). Luckily, the engineering literature may have a solution to our dilemma.

Taguchi methods (Taguchi & Konishi, 1987) can be thought of as a form of non-parametric statistics and research design. That is, there are far less assumptions regarding the distributions of sampling units (sound parameters in our case) in Taguchi methods. Parametric statistics and designs are often praised for their simplicity and accuracy. However, most problems with the conclusions of flawed parameteric studies stem directly from assumption violations (Rosenthal & Rosnow, 2008). Taguchi (1987) asserts that the parameter sampling distribution could be assumed to be random. By using this assumption, it is possible to use a fractional ANOVA structure. That is, each parameter manipulation does not need to be equally presented. Koehl & Parizet (2006) provide an example of an orthogonal display design based on Taguchi methods. The greatest advantage of this sort of research design is that you can get accurate classification of collections of parameters in a minimal amount of tests (presentations in our case). Though, we must assume that there are no interactions between parameters at first. It is then possible to use ANOVA to search for possible interactions by averaging across factors that have one of the levels in common (see Koehl & Parizet 2006). It is now safe to use the assumptions of ANOVA because our assumptions will not be based on the sound per se, but on the average responses to the sounds. So, by using a Taguchi orthogonal array (Taguchi & Konishi, 1987) as a guide for building our stimuli, and our experimental design, we can use ANOVA to analyze a diverse sampling of stimuli. Furthermore, it is likely that some form of a model could be derived from the results, which is probably not the case when using far less stimuli. The next section outlines our proposed parameters by which we can design stimuli according to an orthogonal table.

2.2 Proposed Critical Parameters

2.2.1 Tonal Content

It is suggested that the global spectral content of our sounds contain as little noise as possible. While the state of the art emphasizes noise, it seems likely that such sounds may create many opportunities for confusion and dangerous situations. It is recommended that our stimuli consist of primariy of sinusoidal harmonics. Based on the attention literature, it seems that the harmonics should be between 3-10. However, we are not opposed to using filtered noise as a secondary component to our global structure. It is clear that noise can be very informative. It is also clear that noise tends mask other sounds including noise. Therefore, we propose that any noise components be at least 3 dB less than the frequency components. Furthermore, if noise is used it must be narrow band filtered, which should greatly reduce any masking (if tonal complexity is appropriate) and potential noise pollution. The proposed levels of the ‘tonal content factor are listed below.

Level 1: Simple (3 or less)

Level 2: Moderately Complex (4-6)

Level 3: Complex (7-9)

2.2.2 Frequency Detuning (Modulation)

Level 1: Harmonic

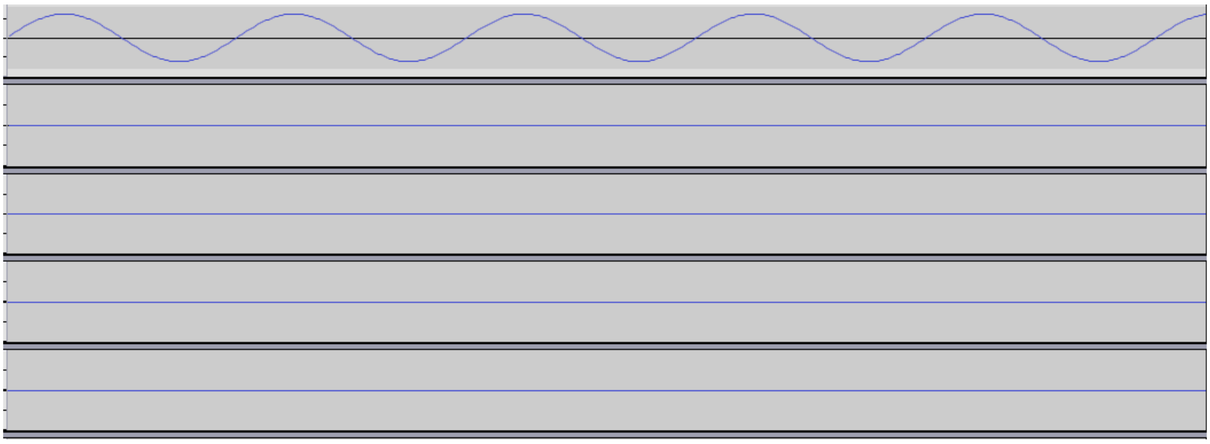
The tonal content within these stimuli will be harmonic. Furthermore, these frequencies will undergo zero frequency modulation for the duration of a trial. For example, a 4 tone harmonic complex with a fundamental frequency (F0) of 300 Hz contains the following tones.

	F0	H1	H2	H3	H4
Hz	300	600	900	1200	1500

Stimuli in the ‘Harmonic’ class will maintain harmony throughout a presentation.

Level 2: Sinususoidal Modulated Detuning

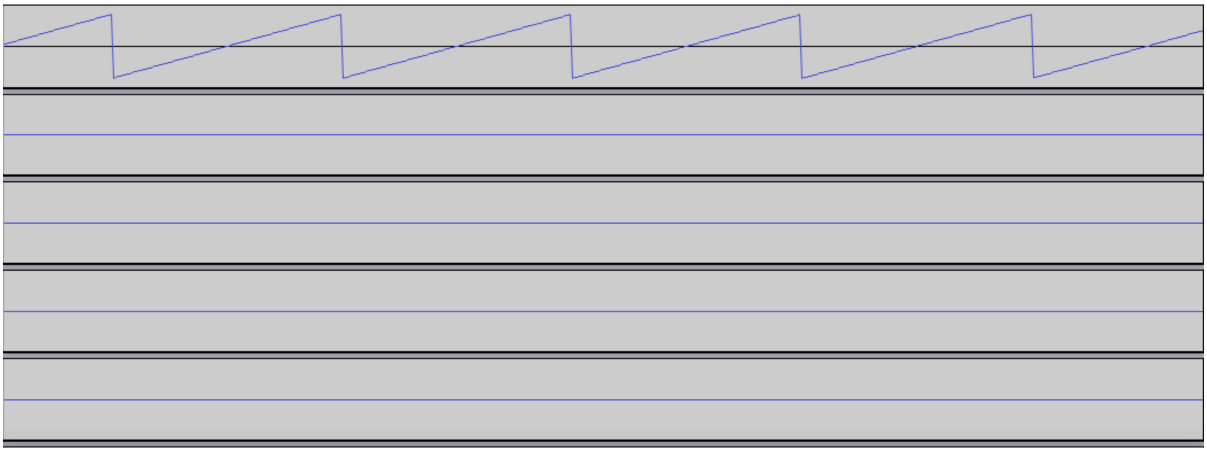
These stimuli will have some harmonics detuned dynamically according to the frequency of a specified sinusoidal oscillation (see fig below). This oscillation will modulate the pitch of the selected frequencies according to a specified percentage of detuning (e.g. +/- 16%). In other words, the pitch of the selected frequencies will constantly and smoothly shift from lower to higher and vice versa.



The 4 flat blue lines in the above figure represent a 4 tone harmonic complex. The fundamental frequency is the lowest line and the 4th line would be the 3rd harmonic. The 5th line represents a 4th harmonic that is detuned according to a sinusoidal function. If the specified percentage of detuning is 16%, the amplitude of the oscillation would be whatever value is specified to be 16% of the fundamental frequency that surrounds (+ or -) the 4th harmonic.

Level 3: Sawtooth Modulated Detuning

The sawtooth modulations in these stimuli are quite different from the sinsusoidal modulations found in the ‘level 2’ stimuli (see figure below).



The primary difference is that the hypotenuse of triangular waveform can either be positive (sweep up), as is demonstrated in the figure above, or negative (sweep down), which is not shown. While the sinusoidal modulations will pass through 0 (harmonic) when going from negative to positive (or vice versa), the sawtooth modulations will be made to start at 0 (harmony), and then either sweep down or up. Furthermore, there are abrupt terminations of the sweeps at the end of each cycle, which certainly sounds distinct from the smooth oscillations and could provide an impulsive sound that could be very useful for listeners.

2.2.3 Amplitude Modulation

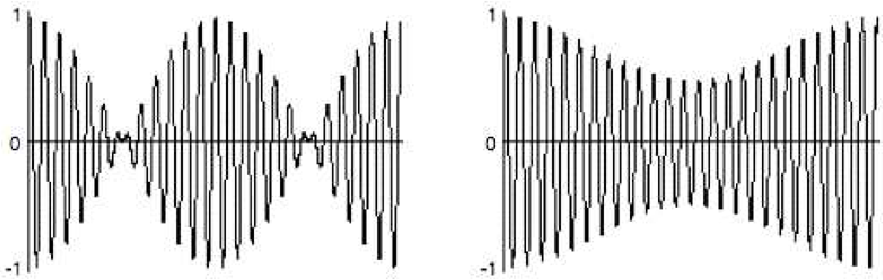
A further suggested departure from the current state of the art is a well-defined amplitude modulation structure. It is proposed that a sound stimulus that has a low frequency amplitude modulation between 1 and 100Hz will create a somewhat sequential and repetitive structure that is similar to an engine. Additionally, the modulation structure can systematically vary such that higher frequency amplitude modulation that could be harmonically to the low frequency modulation. This could be done so that each sound will contain a modulation envelope structure that can produce periodic peaks and valleys that can be informative binaurally. Furthermore, it is proposed that at least some sounds contain sharp peaks in the envelope structure since we know that impulsive sounds are known to be good for detection and can provide rich echoic information. This could also create enough co-modulation for listeners to segregate similar sounds (e.g. other modified vehicles). The levels of the amplitude modulation factor are listed below.

Level 1: Constant (No Amplitude Modulation)

Similar to the ‘Harmonic’ stimuli discussed above, these stimuli will have a constant structure for the duration of a trial. Recall that the harmonic relationship between the tones will remain constant in the ‘Harmonic’ stimuli. In the ‘Constant’ stimuli, the amplitude of the tonal structure will remain constant, as opposed to the harmonic structure.

Level 2: Modulation with Regular Temporal Structure

The parameters for amplitude modulation envelopes for stimuli in this category will be not change for the duration of a presentation (1 trial) (see figure below).



Ring modulation Amplitude modulation

In the above example, the envelop on the left is ring modulated, while the envelope on the right is not. While they obviously look different, and certainly sound very different, both envelopes could be classified as ‘temporally regular’ (AM: level 2 stimuli). They would only be classified as such if the relative gain for each frequency band is maintained throughout the duration of the sound. That is, these types of stimuli contain steady state modulation parameters that may be similar to well-tuned, idling engine. However, the sounds in the above illustration are too simple to make the comparison, and we forecast much more complex modulation patters that may have a pulsating quality that is generated with multiple modulating frequencies. As we know from the research on sequential streaming, such a uniform sequence-like structure is preferable for source segregation and even identification.

Level 3: Modulation with Irregular Temporal Structure(s)

Recall that literature does suggest that some temporal irregularity in a sequential sound can be very useful for attention, learning, perceived urgency, and stream segregation. The Patterson & Mayfield (1990) prototype can be used as an example of such variation (see figure in the Introduction). Notice the variation in the interval between the onsets of the similar pulses. While we cannot be certain if this type of temporal variation will be annoying or even representative of an engine sound, it is not difficult to find these types of variations in ICE vehicles. They are likely realized in the sudden clicks and puffs associated with leaky manifolds and other imperfections. It could be that including such variations may be a way to create very quiet and effective replacement sounds for EVs.

2.3 Proposed Orthogonal Table and Stimuli Selection

The table below is a Taguchi L_9 orthogonal table(see Koehl & Parizet, 2006). This is a guide for the relevant combination of parameters need for an appropriate pseudo-random sample (Taguchi & Konishi 1987).

Orthogonal Table (Taguchi & Konishi, 1987)

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

It should be mentioned that this table is a partial L_9 because we only have 3 factors with 3 levels. The $L_{9\,I}$ is suitable for any experiment with 4 or less parameters (factors) with 3 levels. This means that we have room for one more parameter (e.g. spectral shape after synthesis), but at this stage it is unclear if it is necessary. As you can see, this affords us a much more pragmatic approach to stimuli design, and increases our freedom to manipulate nested factors within these parameters, which I am sure is preferable; especially for concerns of ascetics. It should be noted that the different combinations of parameter manipulations could cause an increase/decrease in the overall loudness of a given sound. It is proposed that the loudness be normalized for all stimuli to prevent confounding results based on loudness.

Practically speaking, these 9 stimuli must be considered as 18 stimuli because it is necessary to present each stimulus as approaching from both the left and the right. Furthermore, as discussed at the April 12 conference in Lyon, it is preferable to have an quasi-random inter-stimulus interval between trials. In that, such a presentation will be more realistic and reduce the likelihood of response bias. It was also established on April 12 that 45 minutes was the maximum amount of time preferred for the actual experiment. Given that we are not planning on truncating the stimulus at response, we can estimate that each trial will take approximately 15-20 seconds. If we limit our the amount of time possible between trials to 20 seconds, then it can be calculated that we can have 2 trials per minute. It follows that we can produce 5 repetitions of each stimulus. I believe this is a suitable number and if the consortium agrees, we can consider this an important milestone in the execution of the eVADER project. Construction of a synthesizer (eVADER synth) has been under way since the April 12th meeting, and is nearly complete. The next section will briefly describe its utility and construction.

3 eVADER Synth

Finally, it is proposed that a custom synthesizer be used to design these sounds using Max/Msp (Chamard et al 2012; Misdariis et al 2012). Such a synthesizer (eVADER synth) has been under construction and is nearly complete. The eVADER synth allows for real-time control of all of the parameters (critical and nested) for an speeded design. Furthermore, it will function like a musical instrument, which should provide more opportunity for the creation of ascetic sounds. As a final point, the eVADER synth will potentially allow us to discover any combination of parameters that may be central to our endeavor. The eVADER synth has been designed to use additive, AM, and FM methods of synthesis. Every manipulation can be stored to be sure any potentially important parameter combinations can be recalled and manipulated later. The infrastructure is most similar to a multichannel studio mixer with multiple options for routing signals through a virtual patch bay. So far, there are 20 sound sources (10 sine wave generators, 10 channels of filtered noise). There are 4 assignable subgroups and every channel can be individual processed by various modules if they are not in a subgroup. While it may seem overly complex, it is flexible so that that it is not necessary for any processing to occur. Once the consortium can agree on parameter selection and methods of stimuli design, the finishing touches will be made on the eVADER synth and it will not be long before we can start testing our human subjects. Some examples of candidate stimuli designed according to methods and parameters described in this proposal will soon be available to provide a more tangible outlook on what can be created.

4 References

Altmann, C.F., Bledowski, C., Kaiser, J., Wibral, M. (2007). Processing of location and pattern changes of natural sounds in the human auditory cortex. *NeuroImage, 35*, 1192-1200.

Best, V., Gallun, F.J., Carlile, S., Shinn-Cunningham, B. (2006). Binaural interference and auditory grouping. *Journal of the Acoustical Society of America,* 1070-1076.

Bregman, A. (1990). *Auditory scene analysis: The perceptual organization of sound.* CMA: The MITP.

Bregman, A. (2004). Auditory Scene Analysis and the Role of Phenomenology in Experimental Psychology. *Canadian Psychology/Psychologie canadienne, 46* (1), 32-40.

Carlyon, R.P. (2004). How the brain separates sounds. *Trends in Cognitive Sciences, 8* (10), 465-471.

Catchpole, K.R., Mckeown, J.D., Withington, D. J. (2004). Localizable auditory warning pulses. *Ergonomics, 47* (7), 748 – 771.

Chamard, J.C. & Roussarie V. (2012). Design of Electric or Hybrid vehicle alert sound system for pedestrians. *Proceedings of the Nantes Acoustics Conference 23-27 April 2012, Nantes, France.*

Edworthy J., Loxley S., Dennis I. (1991). Improving auditory warning design: relationship between warning sound parameters and perceived urgency. *Human Factors , 33* (2), 205-231.

Edworthy, J. & Hards, R. (1997). Learning auditory warnings: The effects of sound type, verbal labeling and imagery on the identification of alarm sounds. *International Journal of Industrial Ergonomics, 24*, 603-618.

Edworthy, J., Hellier, E., Titchener, K., Naweed, A., Roels, R. (2011). Heterogeneity in auditory alarm sets makes them easier to learn. *International Journal of Industrial Ergonomics, 41, 136*-146.

Gaver, W.H. (1993). What in the world do we hear? An ecological approach to auditory event perception. *Ecological Psychology, (5),* 1-29.

Grantham, D.W. (1986). Detection and discrimination of simulated motion of auditory targets in the horizontal plane. *Journal of the Acoustical Society of America, 79*, 1939–1949.

Grantham, D.W., Hornsby, B.W.Y, Erpenbeck, E.A. (2003). Auditory spatial resolution in horizontal, vertical, and diagonal planes. *Journal of the Acoustical Society of America, 114* (2),1009-1023.

Hafter, E.R, Trahiotis, C. (1997) *Functions of the binaural system.* Crocker MJ (ed), Encyclopedia of Acoustics. Wiley.

Kayser, C., Petkov, C.I., Lippert, M., Logothetis, N.K. (2005). Mechanisms for Allocating Auditory Attention: An Auditory Saliency Map. *Current Biology, 15,* 1943–1947.

Kerber S., Fastl H. (2008). Prediction of perceptibility of vehicle exterior noise in background noise. *Proceedings DAGA'08 , Dresden, 2008.*

Koehl, V., Parizet, E. (2006). Influence of structural variability upon sound perception: Usefulness of fractional factorial designs. *Applied Acoustics, 67*, 249–270.

Mcbeath, M.K. & Neuhoff, J.G. (2002). The Doppler effect is not what you think it is: Dramatic pitch change due to dynamic intensity change. *Psychonomic Bulletin & Review, 9* (2), 306-313.

Misdariis, N., Cera, A., Levallois, E., Locqueteau, C. (2012). Do electric cars have to make noise? An emblematic opportunity for designing sounds and soundscapes. *Proceedings of the Nantes Acoustic Conference 23-27 April 2012, Nantes, France.*

Middlebrooks J.C., Green, D.M. (1991) Sound localization by human listeners. *Annual*

Review of Psychology,42, 135–159.
Moore, B.C.J. (2003). *An Introduction to the Psychology of Hearing.* 5th ed. Academic Press, San Diego, CA.
Neuhoff, J.G., Kramer, G., Wayand J. (2002). Pitch and Loudness Interact in Auditory Displays: Can the Data Get Lost in the Map? *Journal of Experimental Psychology. 8*, (2)1, 17–25.
Parizet, E. (2011). Signaux d'alerte : analyse bibliographique. Technical Report to CETIM, 2011.
Patterson, R. D. (I982). Guidelines for auditory warning systems on civil aircraft. *Civil Aviation Authority*. Paper 82017.
Patterson, R.D. (1985). Auditory warning systems for high-workload environments. *Ergonomics International,* 163-165.
Patterson, R.D. (1990). Auditory warning sounds in the work environment. *Philosophical Transcriptions of the Royal Society of London B 327*, 485-492.
Poirson, E., Petiot, J.F., Richard, F. (2010). A method for perceptual evaluation of products by naive subjects: Application to car engine sounds. *International Journal of Industrial Ergonomics, 40*, 504-516.
Razak, K.A. & Fuzessery, Z.M. (2006). Neural Mechanisms Underlying Selectivity for the Rate and Direction of Frequency-Modulated Sweeps in the Auditory Cortex of the Pallid Bat. *Journal of Neurophysiology, 96*, 1303–1319.
Robart, R. & Rosenblum, L.D. (2009). Are hybrid cars too quiet? *Proceedings of the 157th meeting of the Acoustical Society of America:* Paper 5aNS8.
Robart, R. (2010). The effects of spatial and motion properties on auditory grouping. (Dissertation manuscript available upon request).
Rosenthal R., & Rosnow, R.L. (2008). *Essentials of behavioral research: Methods and Data Analysis.* 3rd ed. McGraw-Hill, New York, NY.
Shafer R. M., (1993). *The Soundscape .* Destiny Books.
Stanton, N. & Edworthy, J. (1998). Auditory affordances in the intensive treatment unit. *Applied Ergonomics, 29* (5), 389-394.
Sueid C., Susini P., McAdams S. (2008). Evaluating warning sound urgency with reaction times. *Journal of Experimental Psychology, 14* (3), 201-212.
Sussman, E.S., Bregman, A.S., Wang, W.J., Khan, F.J. (2005). Attentional modulation of electrophysiological activity in auditory cortex for unattended sounds within multistream auditory environments. *Cognitive, Affective, & Behavioral Neuroscience, 5* (1), 93-110.
Tabata, T., Konet, H., Kanuma, T. (2011). Development of Nissan Approaching Vehicle Sound for Pedestrians. *Proceedings of the InterNoise Conference, Osaka Japan.*
Taguchi G, Konishi S. (1987). *Taguchi methods. Orthogonal arrays and linear graphs. Tools for Quality Engineering.* Dearborn (MI): American Supplier Institute Press; 1987.
Treisman, A.M. & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology, 12*, 97–136.
Turgeon, M., Bregman, A.S, Ahad, P.A. (2002). Rhythmic masking release: contribution of cues for perceptual organization to the cross-spectral fusion of concurrent narrow-band noises. *Journal of the Acoustical Society of America, 111*, 1819–1831.
Wang, Y.S. (2008). Sound quality estimation for nonstationary vehicle noises based on discrete wavelet transform. *Journal of Sound and Vibration, 324*, 1124–1140.
Woods, D.L., Alain, C., Diaz, R., Rhodes, D., Ogawa, K.H., (2001). Location and Frequency Cues in Auditory Selective Attention. *Journal of Experimental Psychology: Human Perception and Performance, 27* (1), 65-74.
Zahorik, P., Brungart, D.S., and Bronkhorst, A.W. (2005). Auditory distance perception in humans: A summary of past and present research. *Acta Acustica, 91,* 409-420.