Applied Ergonomics 59 (2017) 153–164

Contents lists available at ScienceDirect

Applied Ergonomics

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

Psychophysical evaluation of auditory signals in passenger vehicles

Chia-Fen Chi*, Ratna Sari Dewi, Min-Hua Huang

Department of Industrial Management, National Taiwan University of Science and Technology, Taipei, Taiwan

ARTICLE INFO

Article history:
Received 20 October 2015
Received in revised form
11 August 2016
Accepted 26 August 2016

Keywords:
Sound quality
Paired comparison
Psychoacoustics

ABSTRACT

Twenty-one experienced drivers were recruited for the evaluation of sounds of four functions (horn, indicator, door open warning, and parking sensor) made by 11 car brand names. Each participant was required to evaluate all of the above sound signals by a pair-comparison test. After the comparison test, each participant was shown his/her pair-comparison result and was asked to comment on their preference and appropriateness of a sound. The physical properties and interview data were compared and summarized to propose design recommendations. Our results indicate that complex tones and a fundamental frequency between 500 and 1000 Hz were most preferred for horns while for indicators the preferred sounds had a higher dominant frequency. To reduce monotony, the indicators with double clicks and an OFF time interval of between 330 and 400 ms between two clicks were most preferred. Regarding door warning sounds, the waveform starting with a higher intensity then fading towards zero intensity is most preferred while for parking sensors, sounds beginning with a longer OFF time (about 500 ms) and having 3 or 4 distinctive tempo variations were most preferred. The relationship between pleasurability and pitch, loudness, and the tempo of sound signals basically followed an inverted-U function. Sound designers should avoid using very extreme parameter values when generating sound for a given function.

© 2016 Elsevier Ltd. All rights reserved.

1. Introduction

The human auditory system is omnidirectional and does not interfere with visual information processing (Sodnik et al., 2008). Auditory alarms were found to induce a greater level of compliance than do visual alarms (Duffy et al., 2004; Wogalter et al., 1993), and professional drivers preferred auditory signals since it could provide a warning as well as arouse the drivers (Meng et al., 2016). Thus, auditory signals are often used to notify drivers about impending danger, operation feedback, and malfunction warnings (Yamauchi et al., 2004) where the operators might risk missing a visual signal (Edworthy, 1994). Several ergonomic considerations of non-verbal auditory warnings relate to the characteristics of a good coding system (Sanders and McCormick, 1993); e.g., detectability, discriminability, meaningfulness. These ergonomic considerations include: (1) The auditory warnings should be reliably audible (Lemaitre et al., 2009) to get one's attention but not too loud to cause startled reactions and impair the primary task performance (Patterson and Mayfield, 1990). (2) A warning should be discriminable based on spectral characteristics such as the fundamental frequency, harmonic series, amplitude envelope shape, speed, rhythm, and melodic structure (Edworthy et al., 1991), and the number of immediate-action warning sounds should be small, not exceeding about five to six (Patterson and Mayfield, 1990). (3) The warning signal should be psychologically appropriate with the signaled situation or have a close signal-referent relationship (Edworthy et al., 2014); in other words, the sound should be meaningful to reduce learning time and promote an immediate and precise response (Patterson and Mayfield, 1990). This third criterion of psychologically appropriateness relates to the sound quality definition "adequacy of a sound in the context of a specific technical goal and/or task," suggested by Blauert (1994).

Sound quality is important for the delivery of customer satisfaction, and as a result, automotive manufacturers began to focus on the sound-quality-related aspects of their vehicles to maintain a competitive advantage (Jennings et al., 2010). In order to design an informative and pleasant sound that supports a positive image of a car (Genuit, 2004), acoustic designers must be aware of how potential customers would react to and appreciate automotive sounds (Otto et al., 1999) and use an appropriate evaluation method to achieve optimum acoustic quality (Humphreys et al., 2011). The perceived acoustic quality is influenced by many attributes, which are divided into three major categories: (1) physical (sound field), (2) psychophysics/psychoacoustic (auditory perception), and (3)

* Corresponding author.
E-mail address: chris@mail.ntust.edu.tw (C.-F. Chi).

http://dx.doi.org/10.1016/j.apergo.2016.08.019
0003-6870/© 2016 Elsevier Ltd. All rights reserved.

psychological (Genuit, 1996; Blauert and Jekosch, 1997). Each of the above categories will be elaborated below.

The physical properties of sound have been characterized by three measureable attributes: frequency, amplitude, and temporal patterns (Yost, 2009), each corresponding to a primary psychoacoustic attributes of sound, e.g., pitch, loudness, and tempo. Besides the above psychophysical attributes, timbre allows one to distinguish among sounds with equal pitch and loudness (ANSI, 1973). It is often referred as "what it sounds like" (Handel, 1995) or "sound color" (Kraus et al., 2009). In terms of psychological dimension, human experience, expectations, and subjective attitudes affect their way in classifying auditory events (Genuit, 2004). Kuwano et al. (2002) have shown that cars with a pleasant sound were perceived to be luxurious. On the contrary, acoustic quality is negative if the auditory signal was perceived as unpleasant, annoying or disturbing (Genuit, 1996). Even though the pleasantness of the product sound is not very high-ranking in the hierarchy of desired features, product sound is still an important factor for customer satisfaction (Blauert and Jekosch, 1997).

Thus, in summary, the identification of sounds results from both a bottom-up and a top-down process (Lemaitre et al., 2009). It is not only what we hear that tells us what we know; what we know also tells us what we hear (Howard and Ballas, 1980). This top-down psychological process helps make the sound signal more resistant to noise and enhance the overall recognition efficiency (Marslen-Wilson and Welsh, 1978). Jennings et al. (2010) applied principal-component analysis to 12 bipolar semantic scales and confirmed that the nature of the sound quality characterization could be reduced to two underlying perceptual dimensions: (1) powerful and (2) refined. However, when different people were asked to assess powerful and refined sound quality dimensions of different vehicles, they might judge different vehicle sounds based on different features of their experience (Jennings et al., 2010). Lemaitre et al. (2009) also found a lack of consensus of listeners in their decision about whether a sound belonged to the category of car horn sounds, and they could even give different responses for sounds rated as being similar.

Since designing correct attributes into a vehicle sound directly impacts the appeal and profitability of the vehicle, and subjective testing is the best way to derive possible attributes affecting sound quality (Otto et al., 1999), the current study collected interior sound signals from horns, indicators, door open warnings, and parking sensors from 11 different vehicle brands and evaluated each with a subjective evaluation test. As humans are only capable of retaining acoustic information for a brief period of time (Bigelow and Poremba, 2014) and perform better in a relative judgment task (Sanders and McCormick, 1993; Huang et al., 2008), the pair-comparison procedure has been adopted for the current evaluation of sound quality. For each sound signal, this study derives important physical attributes (e.g., frequency, amplitude, and temporal pattern) and compares them with post-test interview comments of preferred and non-preferred sounds to understand what factors influence individual appraisals of automobile sounds during the subjective evaluation (Humphreys et al., 2011).

2. Methodology

2.1. Participants and experiment apparatus

Twenty one participants (14 males and 7 females) aged between 29 and 57 (mean = 43.5 years, standard deviation = 8.2 years) took part in this experiment. All participants had a driver license and none had any obvious hearing abnormalities. Prior to the experiment, each participant was briefed about the purpose of the experiment and was given 1 training trial. Each participant was paid the equivalent of 7 US dollars for around 45 min of their participation.

For the experiment, all participants were asked to sit 75 cm in front of two speakers to mimic the sound producing distance in a car. From this distance, the intensity levels of tested sounds measured by TES-1351 Sound Level Meter ranged between 48 and 71 dB(A) depending on the function and the brand. The pair-comparison test module was conducted on one Intel core I7 4770 computer with 8 GRAM and 21.5" VA2248m-LED monitor.

2.2. Paired-comparison test

Overall, each participant was required to accomplish a pairwise comparison test consisting of 86 trials. Each trial was administered with two chosen sound signals presented as two speaker icons which appeared on the left- and right-hand side of the task screen (see Fig. 1). Each participant clicked on these icons to play the corresponding sound signals at least once, and then clicked on one of the circle buttons to indicate his/her preference. However, if the difference between two compared sounds was unnoticeable, the participant could click on "no specific preference" to reduce the chance of circular errors implying that the listeners preferred sound A over B, sound B over C, and sound C over A (Parizet, 2002). The circular error rate for each participant was calculated to identify and exclude non-sensitive participants (Parizet, 2002) and the pair-comparison test was only conducted between sounds of the same function. Humphreys et al. (2011) minimized the car company identity effect by not presenting the identify to the participants as they listened to the tested sound signals. In the end, each participant was shown his/her pair-comparison result. The most and least preferred sounds were replayed for the participants and each participant was asked to explain their reasons for liking and disliking the sounds.

2.3. Collecting sound stimuli

Forty sounds, including 11 horns, 11 indicators, 9 door open

Fig. 1. Pairwise comparison task screen.

warnings, and 9 parking sensors, were recorded from 11 different brands of vehicles (see Table 1). Each sound signal was recorded with a portable recorder placed near the steering wheel position inside the vehicle compartment. According to Kim et al. (2009), sound in a passenger car includes engine, mechanical, and electrical sounds. Our study focused on electrical sounds because contrary to engine and mechanical sounds, electrical sound signals were less affected by human manipulation of vehicles. The indicator, door open warning, and horn were recorded before the car engine was started, while the parking sensor was recorded after turning on the car engine. The brand names of vehicles were chosen by a collaborating car company to serve as the bench mark for their car model. Certain sounds were not collected from some car brands because these brands had no sounds associated with that particular function; e.g., the door open warning for ICO and HEL.

As presented in Table 2, each sound signal could be continuous, intermittent, or a combination of both. All horns and two door open warnings had continuous sounds while indicators and other door open warnings were basically intermittent. Contrary to the indicators, the intermittent door warning sounds had a longer duration with a fading amplitude while the parking sensor sounds had multiple tempos to indicate the distance between the car and other nearby objects. A faster tempo indicated a smaller distance between the car and the objects until the sound became continuous.

The number of paired comparisons (N (N-1)/2) increased by the square of the number of sound stimuli (N) (Otto et al., 1999). Twelve out of 40 collected sound signals were eliminated (denoted by E) because they sounded almost the same as at least one of the other brands. Eventually, 28 sound signals of four different functions were chosen for testing (see Table 1).

Otto et al. (1999) suggested that tested sound signals should be between 3 and 5 s for continuous sounds whereas for the transient sounds, each signal should be presented at least three times (Kuwano et al., 2002). All sound signals tested in our study were edited based on the above suggestions, except parking sensors. For parking sensors that have multiple tempos, each tempo segment was edited into a 2-sec interval and combined together. As shown on the last row of Table 2, the parking sensor is composed of 3 different tempos, each lasting for 2 s.

For the warning sound signals, pitch, intensity, and temporal structure were used to distinguish the perceived urgency (Patterson and Mayfield, 1990). Power spectral analysis was conducted using ChiefSI's FlexDSA to extract up to five peak frequencies for each tested sound. These peak frequency components can be classified into fundamental (f_0) or harmonic, for which harmonic frequency components are integer multiples of the fundamental frequencies (Howard and Angus, 2009).

Table 1
Vehicle sound signals of different brand names and functions collected.

Vehicle brand	Horn	Indicator	Door open warning	Parking sensor
ICO	E	✓	N	E
LU6	✓	E	E	E
LU7	✓	✓	✓	✓
MOU	E	✓	E	✓
HCR	✓	✓	✓	✓
NLI	✓	✓	✓	✓
NMA	E	E	E	✓
TMS	✓	✓	✓	N
HEL	✓	E	N	E
AQ5	✓	✓	✓	✓
TCA	✓	✓	✓	(Interference)

N: No sound associated with function.
E: The sound was screened out because of its similarity with other brands.
✓: The sound was tested in current study.

3. Results

3.1. Pairwise comparison test

For each function, all pair-comparison results constitute a complete evaluating matrix. For each participant, the answer of each pair comparison is denoted as P_{ij}, where i and j stand for the order in row and column, respectively (Huang et al., 2008). If the former sound was preferred to the latter one, P_{ij} equals 1; and vice versa, if the former sound was less preferred, P_{ij} equals -1; if the participant could not decide a preference, P_{ij} equals 0.

Parizet (2002) stated that it can be considered a circular error under the following conditions:

$$\begin{cases} P_{12} \geq A \text{ and } P_{23} \geq A \text{ and } P_{13} \leq -A \text{ or} \\ P_{12} \leq -A \text{ and } P_{23} \leq -A \text{ and } P_{13} \geq A \end{cases} \quad (1)$$

Given that the number of possible triads $A_t^3 = \frac{t!}{3!}$ where t is the number of sound signals, the circular error rate can be derived as in Equation (2):

$$C = \frac{1}{A_t^3} \sum_{1 \leq i,j,k \leq t} \delta_{ijk} \quad (2)$$

where δ_{ijk} can be calculated according to Mao et al. (2004).

$$\begin{cases} \delta_{ijk} = 0 \; while \; P_{ij} + P_{jk} = 0 \\ \delta_{ijk} = Min\{|\mathrm{Max}\left[Min\left(P_{ij} + P_{jk}; 1\right); -1\right] - P_{ik}|; 1\} \; while \; P_{ij} + P_{jk} \neq 0 \end{cases} \quad (3)$$

The value of function δ_{ijk} is limited to 0 and 1 (Mao et al., 2004). In other words, for each response, if there was circular error, $\delta = 1$; if not, then $\delta = 0$. For each warning signal, the circular error rate was calculated for all of the pair-comparison responses. The criterion of $C = 0.25$ circular error rate was adopted to screen out unreliable participants (Mao et al., 2004; Otto et al., 1999). Table 3 indicates that 4 out of 21 participants were excluded from further pair-comparison analysis because they had a greater than 0.25 circular error rate in the parking sensor evaluation. No obvious individual attributes were derived from the four excluded participants; e.g., age, gender, or driving experience.

The pair-comparison result of 17 participants was summarized for these four functions. Table 4 illustrates the pair-comparison result of the horn as an example. One-way analysis of variance (ANOVA) indicates that different brands generated significantly different preference results ($p < 0.01$). Table 5 shows the total comparison ratings of each function averaged by the number of brands so that all preference rating scores can be compared directly. A series of Tukey tests was conducted to categorize the preference score of all vehicle brands into homogeneous subsets indicated by alphabets.

For example, HEL's horn was perceived to be the worst, while HCR's, AQ5's and TMS's were the most preferred. TCA's was perceived as the worst door open warning while LU7's and HCR's were the most preferred. The reason why certain sound signals were preferred over others is revealed from the analyses of the post-test interviews and physical properties of the sounds. The following sections will discuss these results for each function.

3.2. Horn

For each horn sound, the fundamental frequencies and their harmonics were derived from the five peak frequency components

Table 2
Sound waveforms of the tested functions.

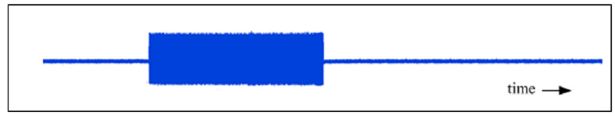
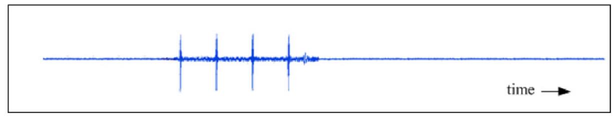
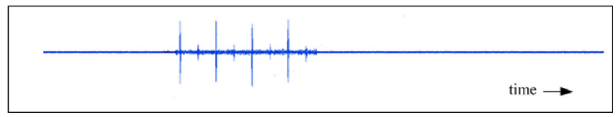
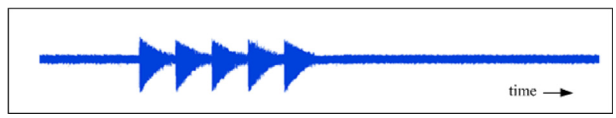
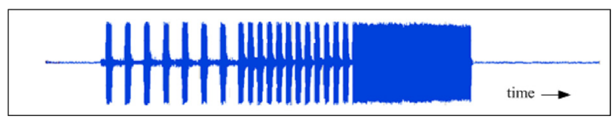
Function	Example of waveforms
Horn Door open warning	
Single-click indicator	
Pair click indicator	
Door open warning	
Parking sensor (with multiple tempos)	

Table 3
The circular error rate of 21 participants.

Participant #	Horn	Indicator	Door open warning	Parking sensor
1	0.005	0.003	0	0.1
2	0.005	0.004	0	0.15
3	0.007	0.005	0.1	0
4	0.008	0.005	0	0.05
5	0.003	0.001	0	0.15
6	0.003	0.002	0	0.1
7	0.005	0.002	0	0
8	0.001	0	0	0.1
9	0	0.001	0	**0.4**
10	0.008	0.005	0	0.1
11	0.005	0.006	0	0.1
12	0.002	0.005	0	0.05
13	0.007	0.01	0	**0.3**
14	0.008	0.007	0.05	**0.3**
15	0.002	0.001	0	0
16	0.005	0.005	0.092	0
17	0.005	0.005	0	0.15
18	0.008	0.007	0.142	0.2
19	0.008	0.008	0	0.05
20	0.006	0.006	0	0
21	0.009	0.007	0.05	**0.267**

Table 4
Pair-comparison results of all 17 participants on horn function.

Participant #	Vehicle brand							
	LU6	LU7	HCR	NLI	TMS	HEL	AQ5	TCA
1	−5	6	2	−6	−3	2	−1	5
2	3	−3	5	−3	3	−7	3	−1
3	1	−1	−1	3	5	−7	3	−3
4	−5	−1	7	−1	−1	−2	−2	5
5	−7	−5	3	−1	7	−1	3	1
6	3	−4	−1	7	5	−6	−1	−3
7	−5	−3	3	1	1	−5	1	7
8	1	−3	1	−5	5	−7	7	1
10	1	−3	5	−5	−1	−3	5	1
11	−5	−3	3	−1	3	−5	7	1
12	−3	−7	3	5	5	−5	3	−1
15	3	5	−5	1	−3	7	−5	−3
16	−4	−4	1	−3	5	−4	6	3
17	−3	5	−1	7	−1	−5	−3	1
18	−5	3	2	1	−5	2	−3	5
19	1	1	1	3	−1	5	−3	−7
20	1	−1	1	1	−3	−7	7	1
Total	−28	−18	29	4	21	−48	27	13

Note that participants #9, #13, #14, and #21 were excluded from the analysis.

using power spectral analysis. HCR's horn, for example, had a fundamental frequency of 450 Hz and a harmonic frequency of 900 Hz; while 366 Hz was identified as another fundamental frequency given that 732 and 3660 Hz were its harmonics. Since the tested horn sounds were continuous, the temporal structure was eliminated.

Based on significant difference in comparison scores using the Tukey test, all horn sounds can be divided into two homogeneous subsets; i.e., A (HCR, AQ5, TMS) and B (HEL) (see Table 5). The other four brands (i.e., TCA, NLI, LU7, and LU6) were listed in both homogeneous subsets (alphabetically coded as AB), meaning each of their preference score was not significantly different from HCR, AQ5, TMS, or HEL. Table 6 summarizes the fundamental frequencies, harmonics, and important comments of 21 participants for horns. The comparison among paired-comparison results, physical attributes, and interview comments indicates that horns with multiple fundamental frequencies (i.e., HCR, AQ5, TMS, and TCA) were preferred over other brands with only one (i.e., NLI, LU7, LU6, and HEL), in particular for a chord sound (more elegant) that has a lower frequency (#10 for TMS). The horn sound has to be loud and clear to generate a warning effect and at the same time not so sharp or noisy that it creates astonishment (#20 for TMS). For HCR and AQ5 horns, a strong and positive familiarity effect was reported by some participants who said the horns sounded similar to their own (#14 for HCR and #8 for AQ5). For horns with a moderate preference score (i.e., TCA, NLI, LU7, and LU6), different participants

Table 5
Mean preference ratings of all four sound signals.

Horn								
Vehicle brand	HCR	AQ5	TMS	TCA	NLI	LU7	LU6	HEL
Mean preference scores	3.63	3.38	2.63	1.63	0.5	−2.25	−3.5	−6
Homogenous subset	A	A	A	AB	AB	AB	AB	B
Indicator								
Vehicle brand	TCA	NLI	AQ5	HCR	MOU	ICO	TMS	LU7
Mean preference scores	8.63	2.63	1.75	1.25	0.75	0.75	−5.75	−10
Homogenous subset	A	AB	AB	ABC	BC	BC	CD	D
Door open warning								
Vehicle brand	LU7	HCR	TMS	NLI	AQ5	TCA		
Mean preference scores	9.5	9.17	−1.17	−4.33	−5.17	−8		
Homogenous subset	A	A	B	BC	BC	C		
Parking sensor								
Vehicle brand	HCR	LU7	NMA	NLI	AQ5	MOU		
Mean preference scores	5.17	3.67	2.33	0.83	0.5	−12.5		
Homogenous subset	A	A	A	A	A	B		

Significant differences in pair comparison score are indicated by alphabetical letters.

Table 6
Fundamental frequencies, harmonics and interview comments on horns.

Vehicle brand	Fundamental frequency (f_0)	Harmonics				Interview result
HCR	450	900				I am used to it as a horn; it has a warning effect (#14).
	366	732	3660			
AQ5	435	870	1305			This horn is not too sharp; the frequency is lower so it is not too noisy (#2).
	520	1560				The horn sound is very similar to the horn of my car; it is loud and clear (#8).
TMS	424	848	1272			Sounds like a chord to make it elegant. The high pitch can increase awareness for warning (#10).
	529	1058				The sound is strong but not sharp; it can warn the driver without astonishment (#20).
						Lack of rich features, it sounds as if it was synthesized from two high-pitch tones (#21).
TCA	431	862	3017			It sounds relatively muffled (#8).
	377	3393				The sound makes people impatient and want to get by quickly (#14).
NLI	360	720	1080	2520	2880	I don't like the sound frequency because it does not have rich features (#21).
LU7	355	1065	1775	2488	2843	The sound matches the expectation for a horn, easy to attract attention (#14).
						The sound is too moderate to have a warning effect (#6).
						The sound pitch is lower than LU6, but the sound is still too sharp (#8).
LU6	418	836	1254	1672	2090	I am used to this sound; it is not too pushy, hasty, or rushed (#3).
						The sound is too sharp (#8).
HEL	372	744	2232	2604	3348	I don't like it because it sounds relatively muffled just like most other horn sounds (#8).
						The sound does not catch people's attention (#12).

Note that for HCR, AQ5, TMS, TCA, the frequency on first fundamental frequency (i.e. 450, 435,424, and 431 Hz) had a higher intensity than the second ones (i.e. 366, 520, 529, and 377 Hz).

either liked or disliked the horns for similar reasons. The preference was associated with familiarity (#14 for LU7 and #3 for LU6). On the other hand, participants disliked sounds that were muffled and lacked a rich feature (#8 for TCA and #21 for NLI), or were too sharp, too pushy, or too hasty that they made people want to get by quickly (#14 for TCA, #8 for LU7, #8 for LU6). For the least preferred horn (i.e., HEL), participants commented that the horn sounded muffled (#8) and did not catch their attention (#12). While some participants preferred a horn sound that was not too sharp or noisy (#2 for AQ5), others suggested that some horn sounds were too moderate to convey a warning effect (#6 for LU7) and that a higher pitch could increase awareness (#10 for TMS).

3.3. Indicator

As opposed to horns, which are continuous sounds, the indicators are intermittent signals with various ON-OFF patterns that can be easily differentiated by a single click or a pair of clicks with different sound signals and OFF time intervals (Fig. 2). Yamauchi et al. (2004) compared the top five stimuli on the suitability with the bottom five among 45 indicators to derive desirable attributes for the indicators. They discovered that the preferred first and second inter-onset interval (IOI) (i.e., a compound of ON time and OFF times) of the top five stimuli ranged from 360 to 400 ms and 330–400 ms, respectively; conversely, the bottom five stimuli ranged from 310 to 340 ms and 310–380 ms. They suggested that the OFF time affected the calmness impression of indicators and that longer OFF time indicators were more suitable and desirable for luxury cars (Yamauchi et al., 2004).

Table 7 presents the ON-OFF intervals, the peak frequency components which were derived from power spectral analysis, and the interview comments of 3 double-click and 5 single-click indicators from most to least preferred in the pair-comparison test. The paired-comparison results indicate that except for NLI, which had a single click and was rated the second-most preferred, all other pair-click indicators had better pair-comparison results (i.e., TCA's, AQ5's and HCR's) than the single-click ones. The most obvious comment that matched with the paired-comparison result was that double-click indicators were less monotonous (#8 for TCA) compared to single-click indicators (#8 for NLI and TMS). Regarding the sound tempo for indicators, participants preferred a tempo that was not too slow (#8 for TCA) and one that had the right

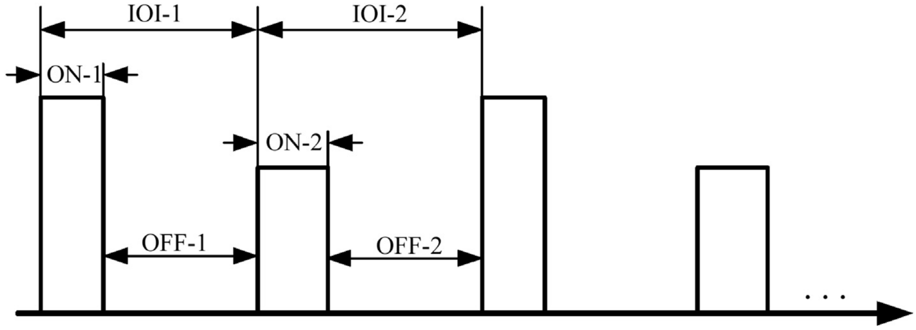


Fig. 2. Pattern diagram of pair click indicators.

Table 7
ON-OFF time intervals (in ms), peak frequencies, and interview comments on indicators.

Brand	ON1 / ON2	OFF1 / OFF2	Peak frequencies (Hz) Click	1st	2nd	3rd	4th	5th	Interview result
TCA	22	328	1st	1858	3282	9342	2625	2749	It sounds like the vehicle that I drive; compared to other tested vehicles, it is smoother, not monotonous; the tempo is not too slow (#8).
	10	339	2nd	1796	2795	1565	1269	9329	
NLI	15	346		2400	5958	3012	1136	2793	I like the sound because the frequency and intensity are appropriate (#12). I don't like it because it is too monotonous compared with the car I drive (#8).
AQ5	20	380	1st	1271	427	1191	629	1009	The sound is not too sharp (#12). Pitch is not too high and the tempo is relatively slow (#21). I have never heard any sound similar to this one before (#8).
	20	380	2nd	833	413	1292	1132	1454	
HCR	16	334	1st	5011	4114	4860	4452	5979	It sounds low and clear (#6). It is not too noisy (#6, 12). Clear, simple, with right tempo (#7). Calm and stable, not too sharp (#21). Frequency and loudness match appropriately (#12). It is not clear and too low. I prefer a higher pitch because it is less likely to be masked by the noise of a moving car (#13). Sounds hasty and not loud enough to be informative (#19).
	16	334	2nd	5755	3379	5945	2688	5083	
MOU	19	341		1885	1977	1414	461	1156	It sounds clear. The pitch is not too high and the tempo is not too fast (#6). Soft, not sharp (#14). It sounds too fast, and makes me anxious and nervous (#7).
ICO	23	337		2484	2296	2074	2756	2989	
TMS	31	648		1718	1880	2069	174	2261	The sound is not too tense, and is loud enough for detection (#2). Stereo sound has enough intensity for aging drivers to notice (#13). The frequency is so high that it is harsh (#7). So monotonous that it makes me feel uncomfortable (#8).
LU7	42	657		1026	936	1135	2002	662	It sounds sharp and a bit blurred (#6). Not solid, not calm, nor clear (#7).

Note that no comment had been given for the indicators of ICO since participants were asked to comment on the most and least preferred.

tempo (#7 for HCR). However, inconsistency was found in the preferred tempo between participants because participant #6 liked MOU for its not-too-fast tempo, while participant #7 thought MOU was too fast, thus causing anxiety and nervousness. Regarding the OFF time, participant #21 did comment that the tempo for AQ5 (OFF time is 380 ms) was relatively slow. Interestingly, no participants commented on the tempo of TMS and LU7 (OFF time = 648 and 657 ms) because human beings tend to focus on the most obvious features. Regarding sound intensity, participants preferred a sound that was not too noisy (#6 and #12 for HCR), was calm, and stable (#21 for HCR), but was also loud enough to notice (#2 for TMS). Regarding pitch, the majority of the participants preferred a sound that was not too sharp (#12 and #21 for AQ5, #21 for HCR, #6 and #14 for MOU) and they disliked a high frequency (#7 for TMS and #6 for LU7). On the contrary, there were also participants who preferred a high pitch because it was less likely to be masked by the noise of a moving car (#13 and #19 for HCR). The appropriate match between pitch and intensity was also emphasized (#12 for NLI and HCR). Lastly, regarding the clear and blurred features of sound timbre, participants preferred sound that was clear (#6 for HCR and MOU), and they disliked sound that was blurred or unclear (#6 and #7 for LU7).

The peak frequency components and ON-OFF pattern for each click sound indicate that for the preferred pair-click indicators, both clicks can be differentiated by either the ON time intervals or the frequency spectrum. The most preferred double-click indicator (i.e., TCA) had significantly different ON times (22 vs 10 ms) between the two clicks, but both clicks had about the same dominant frequency (around 1800 Hz). The second preferred double-click indicator (i.e., AQ5) had about the same ON (and OFF) time intervals between the two clicks, but had a significantly different dominant frequency component between clicks (1271 vs. 833 Hz). On the other hand, the least preferred double-click indicator (i.e., HCR) had no obvious differences in both the ON (and OFF) time intervals or dominant frequency between the two clicks. The above comparison indicates that participants preferred double-click indicators with obvious distinctions between the two clicks either in their temporal patterns or spectral characteristics.

3.4. Door open warning

As shown in Table 5, the paired-comparison results indicate that for door open warnings, participants preferred intermittent sounds (LU7, HCR, TMS, and NLI) over the continuous ones without rhythm (AQ5 and TCA). The most preferred door open warnings (i.e., LU7 and HCR) had a similar waveform in which each signal started with a higher intensity then faded out gradually toward the end of zero intensity, followed by the next repeating signal (see Table 2). Both door open warning signals had a similar dominant fundamental frequency (i.e., 1000 and 801 Hz). On the other hand, two less-preferred intermittent door open warnings (i.e., TMS and NLI) also had similar temporal patterns (i.e., groups of three to four brief duration signals separated by longer pauses as shown in Fig. 3). Between them, NLI had a significantly higher dominant frequency (NLI: 2005 Hz vs. TMS: 872 Hz) and a shorter pause duration (NLI: 536 ms vs. TMS: 1625 ms). Both sounded like an alarm clock, and NLI was even less preferred than TMS due to its high frequency and short pause.

Results from the interview (See Table 8) indicate that participants disliked continuous sounds because they were monotonous (#2 and #16 for AQ5), too sharp, too harsh, or too noisy, and made them feel nervous or agitated (#8, #11, #13, #18, and #19 for AQ5 and #2, #4, #6, #7, #8, #12, #18, #19, #20, and #21 for TCA). However, some participants also suggested that continuous sounds could attract attention and compel the driver to deal with the sound immediately (participant #13 and #14 for AQ5, and #18 for TCA). On the other hand, the most preferred sounds brought a warm, pleasant, and elegant feeling (#10 for LU7, and #1 and #2 for HCR). Participants stated that they preferred LU7 and HCR because they were not too loud or noisy (#1 for LU7 and #6, #19, and #21 for HCR), were slow and calm (#14 for HCR), sounded simple and clear (#12 for HCR), and were familiar (#18 for LU7, and #12 and #18 for HCR). Regarding appropriateness, other than cheap (#2 and #16 for NLI) and too noisy(#16 for NLI), participants either suggested that the sound did not fit in a car (#6 and #8 for TMS) or they identified that the alarm clock sound did not fit in a car (#18, and #21 for TMS; #2, #7, #8, #10, #16, #18, #19, #20, and #21 for NLI in Table 8).

3.5. Parking sensor

As mentioned earlier, all parking sensors were designed to be intermittent with several distinctive tempo patterns to indicate the distance between the vehicle and nearby obstacles, and the fastest tempos always approach a continuous sound. The tested brands had between 2 to an unlimited number of different tempo variations (see Table 9). The paired-comparison results show that MOU and AQ5, which had only two and unlimited tempo variations, respectively, were less preferred than other parking sensors which had three to four tempo variations (see Table 5). For the most preferred parking sensor, HCR, participants indicated that they favored it because it sounded clear (#8 and #20) and not too sharp (#20 and #21), rich (#21), and had distinctive tempo patterns from slow to fast (#20). Similar favorable comments had been given for other parking sensors including clear and informative (#8 for LU7 and #1 for NMA), appropriate intensity and frequency (#9 for LU7 and #6 for AQ5), and urgency (#17 for LU7). On the other hand, the unfavorable comments included that the pitch was too sharp (#6, #14, #18, and #21 for NMA; #21 for NLI and MOU), the tempo was too fast (#7, #14, and #21 for NMA; #7, #8, #12, and #18 for MOU), or both "sharp-pitch" and "fast-tempo" created a tense feeling (#21 for LU7; #6 for NMA, #14 for AQ5). Regarding the appropriateness, similar to the door open warning, parking sensors sounded too much like an alarm clock (#10 for LU7; #2, #12, and #21 for MOU), did not sound like a parking sensor (#14 for MOU), were not distinctive (#10 for LU7) or not informative to tell me how far away the obstacle was (#8 and #10 for MOU). Thus, they were not preferred.

4. Discussion

Theoretically, the optimum sound quality can be developed by an optimum mix of parameter values (Otto et al., 1999). But because of the "Gestalt" phenomena in perception, sound quality results from judgments upon the totality of auditory characteristics of the sound signal. Thus, it becomes difficult to evaluate which attributes of the auditory events play a role in the formation of its sound quality (Blauert and Jekosch, 1997). Our previous study also discovered that a considerable reduction of information takes place from the time of perception to judgment (Chi and Drury, 1998). People may not use all of the relevant information and they tend to use a heuristic rather than analytic approach in reaching a sound quality judgment. Thus, Chi and Drury (1998) used an open-ended question approach to derive what parameters were taken into consideration for the inspection task, and this approach was adopted by our study for our interviewing of the participants.

Blauert and Jekosch (1997) also suggested that the listeners usually reduced the number of parameters necessary to represent the acoustical waves to about less than four to come up with a sound-quality statement. Our interview data suggests that pitch, loudness, the number of fundamental frequencies, tempo variation, and attached meaning, each played a part in participants' perceptions of sound quality. Just like our previous findings (Chi and Drury, 1998), participants realized and commented on the importance of the major parameters, but did not have a clear understanding of how they combined these parameters. Moreover, unlike image, sound is transient; a listener cannot dwell on a sequence of pitches (Miller, 1956). Regarding parking sensors, initially we assumed a greater number of tempo variations was preferred because MOU had only two tempo variations and was significantly less preferred than all the other parking sensors. However, AQ5, which had unlimited tempo variations, was not preferred over other parking sensors with three or four tempo variations. Blattner et al. (1989) suggested that the optimal number of pitches in a motive is between two to four. Thus, with the given limitation in human processing, a parking sensor with a tempo that increased

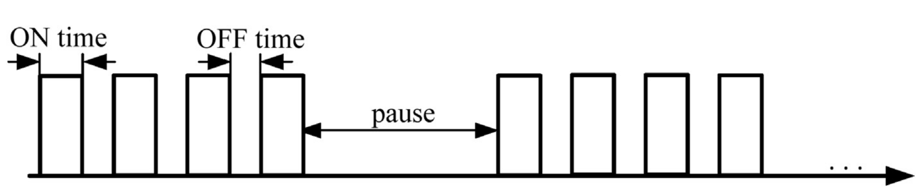


Fig. 3. Pattern diagram of door open warning with group of short signals.

Table 8
ON-OFF time intervals (in ms), peak frequencies, and interview comments on door open warnings.

Brand	ON	OFF	Pause	Peak frequencies (Hz)					Interview result
				1st	2nd	3rd	4th	5th	
LU7	700			1000	2000	3000	177	4000	It sounds soft, not tense or noisy (#1). It sounds familiar (#18). Sounds warm but it also creates a tense feeling for warning (#10).
HCR	983			801	2403	4005	184	167	It brings a pleasant feeling (#1). Sounds like something you heard on an aircraft to give a sense of elegance (#2). Something you heard in a luxury car (#6). It is not too loud for warning (#6, 19). Sounds familiar (#12, 18). Sounds simple and clear (#12). Sounds slow and calm, like the start-up sound of the high-speed rail (#14). The fading pattern is less agitating and noisy (#21).
TMS	97	29	1625	872	585	625	900	851	Sounds like an alarm clock (#18, 21). It does not fit in a car (#8). It does not sound like a door open warning (#6). The high pitch makes me anxious; since it is not urgent, better choose a gentle sound (#6). It is too noisy and loud (#7, 12).
NLI	74	55	536	2005	6015	4010	2064	2147	It sounds like an alarm clock (#2, 7, 8, 10, 16, 18, 19, 20, 21). It does not fit in a car (#8, 20). More related to waking up (#19). Get people a cheap impression (#2, 16). Too noisy (#16). Unpleasant sound (#7).
AQ5				601	1202	5409	1803	3005	Sounds monotonous (#2, 16). Does not fit in a car because high frequency annoys people (#8). Too noisy, harsh, and very annoying (#11, 13, 19). Frequency too high and too sharp (#18). It attracts attention (#13). It affects people's mood (#19). Should have some tempos to make it sound more pleasant (#16). Lasted long to remind people to deal with it immediately (#14). Makes people nervous because it sounds like a patient's heart stopping on a heart monitor (#20). I have heard the sound in a car before but do not like the frequency (#21).
TCA				1952	3904	7808	9760	5856	It creates a sense of warning (#18). I have heard a similar sound in a car before (#21). Prefer not continuous (#4, 6). Feel tense and agitated (#4, 7, 8, 19, 20). Too high frequency (sharp and noisy) (#4, 6, 7, 8, 12, 18, 19, 20, 21). Worse than AQ5 because it is too sharp (#2). Does not fit in a car (#8).

continuously with reduced distance was less preferred.

According to Blauert and Jekosch (2003), sound quality can be improved through objective assessment to find out a better acoustic parameter set. However, due to the nonlinear and non-orthogonal mapping between the acoustic parameter set (frequency, amplitude, and phase) and psychoacoustic dimensions (such as pitch, loudness and timbre), testing sounds created based on a factorial experimental design may be too far from meaningful. Besides, auditory perception can vary with non-acoustic factors, e.g., emotion, knowledge about the situation, and action (Blauert and Jekosch, 2003). Therefore, the testing result of existing sound signals from well-known brands may also have a similar confounding effect as stated in Chi and Dewi (2014).

For an optimal range of acoustic parameters, Berlyne (1971) suggested that the relationship between pleasurability of tones and varying loudness and tempo basically followed an inverted-U function. The inverted-U function concept is found in the majority of our interview comments: participant #2 liked the AQ5 horn because it was not too sharp and not too noisy; participant #20 liked the TMS horn because the sound was strong, but not sharp, and could warn the driver without astonishment (see Table 6); and participant #6 liked the MOU indicator because the pitch was not too high and the tempo was not too fast (see Table 7). Thus, sounds between two extremes are more pleasurable than sounds that are at the extreme of any given parameter. In other words, a very loud, high-pitched sound or a very quiet, low-pitched sound are less pleasurable than sounds within two extremes (Edworthy and Waring, 2006). Although participants were not capable of commenting on all parameters simultaneously, by integrating and comparing analyses from our paired-comparison results, interview data, important sound attributes, and our literature review, we could derive an optimal range for important acoustic parameters.

For intensity level, Patterson (1982) recommended a minimum level of 15 dB and a maximum 25 dB above the masked threshold. Deatherage (1972) suggested that as a rule of thumb, the optimum signal intensity should be about midway between the masked threshold of the signal in the presence of noise and 110 dB. Most importantly, the sound designer should always conform to existing standards. For example, the ECE-R28 agreement from The United Nations Economic Commission for Europe (UNECE) (1972) suggested a range from 93 to 112 dB(A) for car horns, and ISO 7731 (2003) suggested a range from 65 to 118 dB(A) for auditory warning signals. Car manufacturing had suggested a feasible intensity level based on the development of Nissan's approaching-vehicle sound for pedestrians (VSP) used in electric vehicles. The forward sound pressure level is 55 dB(A) to provide detectability for pedestrians and to maintain a quiet environment for drivers and neighborhoods (Tabata et al., 2011).

Regarding pitch, Patterson (1982) suggested that the signal should be composed of four or more dominant frequency components in the range from 1000 to 4000 Hz because complex sounds are more difficult to mask than simpler sounds. A more specific

Table 9
Tempo variations, ON-OFF time intervals (in ms), peak frequencies, and interview comments on parking sensors.

Brand	# of tempo variation	ON-1 ON-2 ON-3	OFF-1 OFF-2 OFF-3	Peak frequencies (Hz)					Interview result
				1st	2nd	3rd	4th	5th	
HCR	4	117 117 117	493 224 40	745	2235	5215	536	3725	It sounds very clear (#8, 20). Not so sharp to cause a tense feeling (#20, 21). Distinctive rhythm patterns from slow to fast (#20). Sounds relatively rich (#21).
LU7	3	93 93	199 62	1993	3986	1969	1889	2068	Sounds very clear, it informs the distance between vehicle and the obstacle (#8). Has appropriate intensity and freq. not too loud (#9). The intermittent pause of first tempo pattern is not too urgent or tense (#17). Sounds like an alarm clock, monotonous, not distinctive (#10). The frequency and tempo make people nervous (#21).
NMA	3	75 75	225 75	2563	2592	2734	2534	2613	Has very distinctive tempo and sound intensity (#1). The frequency is too sharp (#6, 14, 21). It makes people nervous (#6). The frequency of the last continuous sound is too high (#18). The tempo is too fast (#7, 14, 21).
NLI	3	175 119	71 14	2775	2791	2760	2808	5550	I do not like it because it sounds too fast; better to prolong the pause between sounds (#7). Too sharp and the tempo is too fast (#21).
AQ5[a]	∞	101	118	749	2247	5243	1498	3745	I like the appropriate freq. and intensity for warning the driver (#6). I don't like it because it creates a tense feeling (#14).
MOU	2	93	35	3684	3649	3860	3743	3804	Sounds like an alarm clock (#2, 12, 21). The sound is too fast (#7, 8, 12, 18). The intensity is OK (#7). The sound is too quiet (#18). The last tempo pattern sounds too sharp (#21). The sound does not tell me how far away the obstacle is (#8, 10). Warns that the car is about to hit something (#2). Compared with other brands, it does not sound like a parking sensor (#14).

[a] The number of tempo variation for AQ5's parking sensor was unlimited as the tempo was getting faster continuously. Table 9 only presents the ON-OFF time interval of AQ5's first tempo variation.

optimum range for each tested function should depend on the notation of "compatibility". Compatibility was defined as the "adequacy of a sound in the context of a specific technical goal and/or task" to be incorporated as one important attribute for sound quality (Blauert, 1994). From the interview data, our results reveal that "goodness of fit", including perceived urgency and attached meaning (Blauert and Jekosch, 2003), can have a significant effect on the perceived sound quality. Perceived urgency is known to increase with tempo and loudness. If matching urgency to the conveyed message can help to design unobtrusive auditory interfaces (Gaver, 1997), then the perceived urgency of a parking sensor should match with the distance between the car and the nearby objects. In this way, the urgency of graded fatigue warning system proposed by Meng et al. (2016) should match with different danger levels.

Besides perceived urgency, familiarity also has a significant impact on subjective preference (North and Hargreaves, 1995). Participants evaluated the sounds differently based on what type of car they were driving (Humphreys et al., 2011; Jennings et al., 2010). However, familiarity serves as a double-edged sword. Participants preferred sounds that could be associated with something expensive or elegant; for example, the door open warning of HCR sounded like something you would hear in an aircraft (#2). On the other hand, participants disliked sound signals that could be associated with something cheap; e.g., an alarm clock (#18 and #21 for TMS and #2, #7, #8, #10, #16, #18, #19, #20, and #21 for NLI) or something unfortunate, e.g., the long-beep sound of AQ5 sounded like a heart monitor when the patient's heart stops beating (#20). Lemaitre et al. (2009) also suggested that when introducing new warning signals, care must be given to make sure that these sounds are not too different from the existing ones. The more the new signals are different from the existing ones, the more the road users will need time to learn their meanings. Thus, for horns, the chosen fundamental frequency should be around 440 Hz to be a compatible match with the car horns sold in Europe (Ballas, 1993). Our finding that the fundamental frequency of 450 Hz was preferred for a horn matches that suggested range.

Cooper (1977) reported that continuous loud sounds tend to have a detrimental effect on the pilot and crew, and previous research also suggested the use of intermittent signals to reduce potential masking (Doll and Folds, 1986) and perceptual adaptation (Sanders and McCormick, 1993, pp 176). In order to minimize perceptual adaptation, whenever feasible, steady-state signals should be avoided, and interrupted or variable signals should be used (Mudd, 1961; Licklider, 1961). This agrees with our findings that participants preferred less monotonous, complex sounds for horns, intermittent sounds over continuous sounds for door open warnings, and double-click over single-click indicators. In our experiment, all horn sounds were edited into 3-sec intervals for better comparison. But, in fact, the driver can improvise the temporal pattern of a horn sound depending on the perceived urgency of the driving situation or the aggressiveness of the driver (Shinar and Compton, 2004).

As opposed to the horn sound, the three other tested functions (indicators, door open warnings and parking sensors) each had different temporal patterns with the potential to become the most important parameter influencing the participants' comparison decision. Regarding OFF times of the indicators, our pair-comparison results indicate that single-click indicators (TMS and LU7) with significantly longer OFF time (648 and 657 ms) were less preferred than other indicators with a shorter OFF time (328–380 ms). The contradictory finding with Yamauchi et al. (2004) could be due to the fact that Yamauchi et al. (2004) only tested double-click indicators that had an OFF time interval from 310 to 400 ms. Previous ergonomic design guidelines provided a maximum possible range of between 100 and 2000 ms cycle time (Campbell et al., 2004), and the exact range must be fine-tuned based on experimental study in real context (Sanders and McCormick, 1993, pp. 176). Regarding the duration of a signal (i.e., ON time), Gales (1979) suggested that discrimination is best when stimulus duration is in excess of 0.1 s. But a brief duration signal, if combined with a higher frequency pitch, can still be audible (Doughty and Garner, 1947, 1948). Doughty and Garner (1947, 1948) found that for very short tones, like clicks, the duration threshold was about 4–5 ms for high-pitched clicks with a greater than 500 Hz frequency. Based on our study results, the double-click indicators with easily differentiated ON-OFF patterns between clicks and a dominant frequency between 1000 and 2000 Hz were more preferred. Based on the most preferred clicks of between 10 and 22 ms ON time and the repetition frequency range of between 0.5 Hz and 4 Hz suggested by ISO 7731 (2003), the OFF time duration should be between 330 and 400 ms.

Regarding door open warnings, our study found that intermittent sounds with a fading intensity waveform and a dominant fundamental frequency of between 500 and 1000 Hz were the most preferred, which agrees with those suggested in Deatherage (1972). As for the parking sensors, the most preferred sounds had a dominant fundamental frequency between 500 and 2000 Hz, 3 to 4 distinctive tempo variations, and started at a longer OFF time pattern, e.g., around 500 ms. Regarding the favorability of all of the above functions, other than richness of sound, sound clarity is another important attribute for sound quality based on participants' comments stating that they preferred sounds that were clear, not blurred or muffled (#8 for AQ5 horn, #6 and #7 for HCR and #6 for MOU indicators; #12 for HCR open door warning, #8 and #20 for HCR parking sensors).

In addition to the discussion about the sound signal, another interesting observation about the experiment is that one of the participants with a PhD degree in Industrial Design and Ergonomics was excluded due to circular error. Weber (1999) suggested three possible explanations for participants' misjudging: (1) not enough concentration during the evaluation process, (2) vacillation of judging criterion, and (3) little difference between sound events. For our experiment, the excluded participant explained during the interview that each pair of sounds differed in more than one aspect; thus, she had a difficult time in trying to choose which attribute (loudness, frequency, or timbre) influenced her preference. She also revealed that the decision was difficult at times when she preferred one signal for its frequency and the other for its intensity. Just as stated by Blauert and Jekosch (1997), representative listeners are not necessarily expert listeners and to select representative listeners for the evaluation procedures is a complex task in itself. The reason why sound quality evaluation does not require a high level of expertise could be because experts tend to be too analytical to get the point of judging sound quality (Blauert and Jekosch, 2003). Another possible source of circular error is a possible sequential effect (Jesteadt et al., 1977; Chi and Drury, 1988) in the paired-comparison. Even though pairwise comparison is known to be inefficient, it is more appropriate than other psychological methods since human beings can only retain acoustic information for a relatively brief period of time (Bigelow and Poremba, 2014). In order to minimize the potential sequential effect or play-order effect, the pair-comparison test was self-paced by the participants meaning that the user had control of the test and could play the sounds as many times as necessary (Otto et al., 1999). Perhaps with proper calibration and training, dichotic listening is a possible approach to minimize the sequential effect and reduce the pair-comparison time by half.

Importantly, our results are about preference evaluation not performance. There are possible inconsistencies between detection performance and preference evaluation and among the preference opinions collected from different participants or tested under different context. Participants preferred sounds that were soft (Meng et al., 2016), inconspicuous (Genuit and Fiebig, 2014), intermittent, rich and with appropriate tempo and pitch (current study). However, when signal detection was taken into consideration, the participants' preference changed. For example, Antin et al. (1991) collected the minimum intensity level required to achieve a consistent criterion detection rate as well as the preferred tone intensity level under different driving speeds or radio conditions. Surprisingly, the preferred intensity levels were louder than those required to achieve the detection criterion under different driving speeds. Interestingly, when tested in the radio condition, participants preferred lower intensity level in order to minimize broadcast disturbance (Antin et al., 1991). In addition to the inconsistency between signal detection and preference, Meng et al. (2016) also discovered inconsistent opinions among participants about whether the fatigue warnings should be designed into graded or single stimuli. Even though participants agreed that a sound of a fire alarm could be annoying, it did cause them to be alert (Meng et al., 2016).

Contrary to our previous study in which ISO-adopted icons had a much better matching performance than the non-ISO adopted ones, there is no standardized ISO sound signals for vehicles. The existing ISO standard for auditory signals are mostly design principles, such as ISO 7731 (2003), which suggested that a warning signal should include frequency components within the 500–2500 Hz range, more specifically, two dominant components from 500 to 1500 Hz are recommended. There are existing standards for alarms, such as those specified in International Medical Equipment (IEC 60601-1-8). However, Edworthy et al. (2014) tested a set of IEC tonal alarms and compared them to a set of indirect metaphor icons to prove that existing IEC alarms are very difficult to learn while the indirect metaphorical icons can be learned almost instantly due to the closeness of the signal reference relationship and acoustic variability.

Thus, in summary, our research findings of the preferred sounds for different functions can be used as a reference frame so the sound designers may have a sensible start with a feasible range on limited acoustic parameter combinations. In this way, sound designers can design mild warning signals with minimum annoyance based on our preference findings, then increase the intensity or frequency to map with the urgency of a situation. Our current research indicates that no single automobile company had the most preferred auditory signals across different functions. It is very likely that the automobile industry will develop standardized auditory signals similar to the ISO standard icons in the future. The systematic process for developing in-vehicle icons proposed by Campbell et al. (2004), which incorporated preference (ranking test) and recognition performance, can be replaced with a paired-comparison test and a signal-detection test, respectively, for the development of auditory signals to be used in electric vehicles (Genuit and Fiebig, 2014). Our paired-comparison test approach can be adopted as part of the process for the development and evaluation of auditory signals.

5. Conclusion

New in-car technologies have led to an increasing number of sound signals (Suied et al., 2008). In order to reduce the confusability of a warning set, the number of immediate-action warning sounds should not exceed about five to six, each with a distinctive melody and temporal pattern (Patterson and Mayfield, 1990). The current study tested sound signals for horns, indicators, door open warnings, and parking sensors, from 11 car brands, using a paired-comparison test and an interview to gain more insight into human preference of sounds. By comparing analyses from our paired-comparison results, interview data, important sound attributes, and our literature review of existing standards, we propose an optimal range for important acoustic parameters. Sound designers can produce sounds within a feasible range, based on available standards. Then, they can choose a dominant fundamental frequency between 500 and 2000 Hz and adjust based on best practices; e.g., a lower fundamental frequency range for horns (between 440 and 480 Hz) and a higher dominant frequency for indicators (up to 2400 Hz). Since the relationship between pleasurability of tones and a given parameter basically followed an inverted-U function, sound designers should avoid using very extreme parameter values. Lastly, warnings should be readily distinguishable among themselves in terms of temporal pattern (Patterson, 1982), ON/OFF ratio, and attenuation pattern (Meredith and Edworthy, 1994) to reduce monotony and further improve the perceived quality.

References

ANSI, 1973. American National Psychoacoustical terminology.S3.20.-1973 New York. American National Standards Institute.

Antin, J.F., Lauretta, D.J., Wolf, L.D., 1991. The intensity of auditory warning tones in the automobile environment: detection and preference evaluations. Appl. Ergon. 22, 13–19.

Ballas, J.A., 1993. Common factors in the identification of an assortment of brief everyday sounds. J. Exp. Psychol. Hum. Percept. Perform. 19 (2), 250.

Berlyne, D.E., 1971. Aesthetics and Psychobiology. Appleton-Century-Crofts, New York.

Bigelow, J., Poremba, A., 2014. Achilles' ear? Inferior human short-term and recognition memory in the auditory modality. PloS One 9 (2), e89914.

Blattner, M.M., Sumikawa, D.A., Greenberg, R.M., 1989. Earcons and icons: their structure and common design principles. Human–Computer Interact. 4 (1), 11–44.

Blauert, J., 1994. Product-sound design and assessment: an enigmatic issue from the point of view of engineering?. In: INTER-NOISE and NOISE-CON Congress and Conference Proceedings (Vol. 1994, No. 5, Pp. 857–862). Institute of Noise Control Engineering.

Blauert, J., Jekosch, U., 1997. Sound-quality evaluation–a multi-layered problem. Acta acustica united acustica 83 (5), 747–753.

Blauert, J., Jekosch, U., 2003. Concepts behind sound quality: some basic considerations. In: INCE Conference Proceedings (Vol. 206, No. 1, Pp. 72–79).

Campbell, J.L., Richman, J.B., Carney, C., Lee, J.D., 2004. In-vehicle Display Icons and Other Information Elements Volume 1. Guidelines (Rep. No. FHWA-RD-03–065). Federal Highway Administration, McLean, VA.

Chi, C.F., Dewi, R.S., 2014. Matching performance of vehicle icons in graphical and textual formats. Appl. Ergon. 45 (4), 904–916.

Chi, C.F., Drury, C.G., 1988. A further note on psychophysical testing of handles. Appl. Ergon. 19 (4), 315–318.

Chi, C.F., Drury, C.G., 1998. Do people choose an optimal response criterion in an inspection task? IIE Trans. 30 (3), 257–266.

Cooper, G.E., 1977. A Survey of the Status of and Philosophies Relating to Cockpit Warning Systems (No. NASA-CR-152071).

Deatherage, B.H., 1972. Auditory and other sensory forms of information presentation. In: Van Cott, H.P., Kinkade, R.G. (Eds.), Human Engineering Guide to Equipment Design. Government Printing Office, Washington, DC, pp. 123–160.

Doll, T.J., Folds, D.J., 1986. Auditory signals in military aircraft: ergonomics principles vs. practice. Appl. Ergon. 17, 257–264.

Doughty, J.M., Garner, W.R., 1947. Pitch characteristics of short tones. I. Two kinds of pitch threshold. J. Exp. Psychol. 37 (4), 351.

Doughty, J.M., Garner, W.R., 1948. Pitch characteristics of short tones. II. Pitch as a function of tonal duration. J. Exp. Psychol. 38 (4), 478.

Duffy, V.G., Ng, P.P., Ramakrishnan, A., 2004. Impact of a simulated accident in virtual training on decision-making performance. Int. J. Ind. Ergon. 34 (4), 335–348.

Edworthy, J., 1994. The design and implementation of non-verbal auditory warnings. Appl. Ergon. 25 (4), 202–210.

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

Edworthy, J., Waring, H., 2006. The effects of music tempo and loudness level on treadmill exercise. Ergonomics 49 (15), 1597–1610.

Edworthy, J., Page, R., Hibbard, A., Kyle, S., Ratnage, P., Claydon, S., 2014. Learning three sets of alarms for the same medical functions: a perspective on the difficulty of learning alarms specified in an international standard. Appl. Ergon. 45, 1291–1296.

Gales, R., 1979. Hearing characteristics. In: Harris, C.M. (Ed.), Handbook of Noise Control, second ed. McGraw-Hill, New York.

Gaver, W.W., 1997. Auditory Interfaces. Handbook of Human-Computer Interaction, vol. 1, pp. 1003–1041.

Genuit, K., 1996. Objective evaluation of acoustic-quality based on a Relative Approach. In: Inter-noise 1996. Proceedings, Liverpool, England.

Genuit, K., 2004. The sound quality of vehicle interior noise: a challenge for the NVH-engineers. Int. J. Veh. Noise Vib. 1 (1–2), 158–168.

Genuit, K., Fiebig, A., 2014. Sound design of electric vehicles - challenges and risks. Inter-noise 2014. Page 1 of 10. Melbourne.

Handel, S., 1995. Timbre perception and auditory object identification. In: Moore, B.C.J. (Ed.), Hearing: Handbook of Perception and Cognition, second ed. Academic Press, San Diego, London, pp. 425–461.

Howard, D.M., Angus, J.A.S., 2009. Acoustics and Psychoacoustics, fourth ed. Focal Press, Oxford.

Howard, J.H., Ballas, J.A., 1980. Syntactic and semantic factors in the classification of nonspeech transient patterns. Percept. Psychophys. 28 (5), 431–439.

Huang, Y., Di, G., Zhu, Y., Hong, Y., Zhang, B., 2008. Pair-wise comparison experiment on subjective annoyance rating of noise samples with different frequency spectrums but same A-weighted level. Appl. Acoust. 69 (12), 1205–1211.

Humphreys, L., Giudice, S., Jennings, P., Cain, R., Song, W., Dunne, G., 2011. The influence of company identity on the perception of vehicle sounds. Ergonomics 54, 338–346.

International Organization for Standardization (ISO), 2003. Ergonomics. Danger Signals for Public and Work Areas (Standard No. ISO 7731:2003). ISO, Geneva, Switzerland.

Jennings, P.A., Dunne, G., Williams, R., Giudice, S., 2010. Tools and techniques for understanding the fundamentals of automotive sound quality. Trans. IMechE Part D J. Automob. Eng. 224 (10), 1263–1278.

Jesteadt, W., Luce, R.D., Green, D.M., 1977. Sequential effects in judgments of loudness. J. Exp. Psychol. Hum. Percept. Perform. 3 (1), 92.

Kim, T.G., Lee, S.K., Lee, H.H., 2009. Characterization and quantification of luxury sound quality in premium-class passenger cars. Proc. Institution Mech. Eng. Part D J. Automob. Eng. 223 (3), 343–353.

Kraus, N., Skoe, E., Parbery-Clark, A., Ashley, R., 2009. Experience-induced malleability in neural encoding of pitch, timbre, and timing. Ann. N. Y. Acad. Sci. 1169 (1), 543–557.

Kuwano, S., Fastl, H., Namba, S., Nakamura, S., Uchida, H., 2002. Subjective evaluation of car door sound. Proc. Sound-Quality Sym., Dearborn MI.

Lemaitre, G., Susini, P., Winsberg, S., McAdams, S., Letinturier, B., 2009. The sound quality of car horns: designing new representative sounds. Acta acustica united Acustica 95 (2), 356–372.

Licklider, J.C.R., 1961. Audio Warning Signals for Air Force Weapon Systems. Behavioral Sciences Laboratory, Aerospace Medical Laboratory, Wright Air Development Division, Air Research and Development Command, United States Air Force.

Mao, D., Yu, W., Gao, Y., Wang, Z., 2004. Error rating of paired comparison listening test for sound quality assessment. Proceeding ICA 2004.

Marslen-Wilson, W.D., Welsh, A., 1978. Processing interactions and lexical access during word recognition in continuous speech. Cogn. Psychol. 10 (1), 29–63.

Meng, F., Li, S., Cao, L., Peng, Q., Li, M., Wang, C., Zhang, W., 2016. Designing fatigue warning systems: the perspective of professional drivers. Appl. Ergon. 53, 122–130.

Meredith, C., Edworthy, J., 1994. Sources of confusion in intensive therapy unit alarms. In: Stanton, N. (Ed.), Human Factors in Alarm Design. Taylor & Francis, Inc, Bristol, PA, pp. 207–219.

Miller, G.A., 1956. The magical number seven, plus or minus two: some limits on our capacity for processing information. Psychol. Rev. 63 (2), 81.

Mudd, S.A., 1961. The Scaling and Experimental Investigation of Four Dimensions of Pure Tone and Their Use in an Audio-visual Monitoring Problem (Doctoral dissertation, Purdue University).

North, A.C., Hargreaves, D.J., 1995. Subjective complexity, familiarity, and liking for popular music. Psychomusicology A J. Res. Music Cognition 14 (1–2), 77.

Otto, N., Amman, S., Eaton, C., Lake, S., 1999. Guidelines for Jury Evaluations of Automotive Sounds (No. 1999-01-1822) (SAE Technical Paper).

Parizet, E., 2002. Paired comparison listening tests and circular error rates. Acta acustica united acustica 88 (4), 594–598.

Patterson, R.D., 1982. Guidelines for Auditory Warning Systems on Civil Aircraft. Civil Aviation Authority.

Patterson, R.D., Mayfield, T.F., 1990. Auditory warning sounds in the work environment [and Discussion]. Philosophical Trans. R. Soc. Lond. B Biol. Sci. 327 (1241), 485–492.

Sanders, M.S., McCormick, E.J., 1993. Human Factors in Engineering and Design, seventh ed. McGraw-Hill, Singapore.

Shinar, D., Compton, R., 2004. Aggressive driving: an observational study of driver, vehicle, and situational variables. Accid. Analysis Prev. 36 (3), 429–437.

Sodnik, J., Dicke, C., Tomažič, S., Billinghurst, M., 2008. A user study of auditory versus visual interfaces for use while driving. Int. J. Human-Computer Stud. 66 (5), 318–332.

Suied, C., Susini, P., McAdams, S., 2008. Evaluating warning sound urgency with

reaction times. J. Exp. Psychol. Appl. 14 (3), 201.
Tabata, T., Konet, H., Kanuma, T., 2011. Development of Nissan Approaching Vehicle Sound for Pedestrians: How to Solve the Trade off between Quietness and Pedestrian Safety of the Electric Vehicles? (No. 2011-39-7231) (SAE Technical Paper).
United Nations Economic Commission for Europe (UNECE), 1972. Regulation No. 28 Uniform Provisions Concerning the Approval of Audible Warning Devices and of Motor Vehicles with Regard to Their Audible Signals. Retrieved from. http://www.unece.org/fileadmin/DAM/trans/main/wp29/wp29regs/r028e.pdf.
Weber, R., 1999. Interior car Sound Quality—assessment of Acceleration Noises, 2nd Forum Acusticum. Berlin, Germany.
Wogalter, M.S., Kalsher, M.J., Racicot, B.M., 1993. Behavioral compliance with warnings: effects of voice, context, and location. Saf. Sci. 16 (5), 637–654.
Yamauchi, K., Choi, J.D., Maiguma, R., Takada, M., Iwamiya, S.I., 2004. A basic study on universal design of auditory signals in automobiles. J. physiological Anthropol. Appl. Hum. Sci. 23 (6), 295–298.
Yost, W.A., 2009. Pitch perception. Atten. Percept. Psychophys. 71 (8), 1701–1715.