An exploratory study on proposed new sounds for future products

Youyi Bi[a)], Tahira Reid[b)] and Patricia Davies[c)]

(Received: 11 November 2016; Revised: 31 March 2017; Accepted: 24 May 2017)

Car designers are interested in understanding what attributes of naturally occurring, generated and modified sounds make them more or less desirable to end-users. In this research, we investigated millennials' perception of proposed next-generation car sounds and other product sounds. An auditory perceptional test was conducted to determine sound preferences in which people were presented with a current sound, a very different sound and something in between the two. Intentional sounds and consequential sounds were considered in six contexts. Because of the focus on next-generation cars, responses from millennials are of particular interest. The very different sounds were inspired by the musical preferences of the millennial generation. The influence of visual information and perceived functionality on the sound preferences was also examined. Forty university students and staff volunteered to participate in the test. The results showed that millennials preferred traditional sounds in most contexts and their sound preferences aligned with certain sound evaluations and verbal descriptions. Participants' verbal descriptions of the sounds provided interesting insights into the relationship between the sound evaluations and participants' perception of the sounds. In several cases, the order in which the pictorial and textual cues of context were presented impacted how people rated the sounds. These results may shed light on how to integrate millennials' preferences into the design of future product sounds.

Primary subject classification: 63.7; Secondary subject classification: 79.2

1 INTRODUCTION

Customers utilize information from various sources to evaluate products, including the product sounds[1]. Interior and exterior car sounds have been found to influence customers' holistic evaluation of vehicles such as whether a car is luxurious or cheap, sporty or economical, etc. For example, Filippou et al.[2] found that, if the car door sound is tinny, the impression is that the whole vehicle is cheap and not solid. In contrast, a full saturated sound when closing a car door gives the impression of luxury. Parizet et al.[3] noticed that, while hearing the noise coming from a car's door closure, two timbre parameters of the sound, the frequency balance and its cleanness, are of the greatest importance to listeners' evaluation of the quality of the car. Hyeonho et al.[4] found that that the quieter and less sharp the sound, the more luxurious the quality. Pick et al.[5] stated that consumers' initial reactions to a new technology like replacing a conventional transmission with a continuously variable transmission (CVT) may be negative. One reason for this is that the discrepancy between initial engine rpm and vehicle speed is not consistent with people's prior experience with conventional transmissions. These studies show that product sounds can influence how customers evaluate products and new technologies.

To improve the quality of the sounds produced in cars, automotive companies are interested in understanding what attributes of naturally occurring, generated, and modified sounds make them more or less desirable to drivers and passengers. This study was conducted to investigate how millennials would evaluate artificially created sounds of cars and other products. The results from this study are a first step in developing an understanding of millennials' preferences for product sounds.

Among the purchasers of next-generation cars, millennials (born in 1980s to around 2000) have massive spending power. According to a 2016 survey on the millennial generation's spending habits from Gallup[6],

a) The REID Lab, School of Mechanical Engineering, Purdue University, 585 Purdue Hall, West Lafayette, IN 47907, USA and Ray W. Herrick Laboratories, School of Mechanical Engineering, Purdue University, 177 South Russell Street, West Lafayette, IN 47907-2099, USA; email: bi9@purdue.edu.
b) The REID Lab, School of Mechanical Engineering, Purdue University, 585 Purdue Hall, West Lafayette, IN 47907, USA.
c) Ray W. Herrick Laboratories, School of Mechanical Engineering, Purdue University, 177 South Russell Street, West Lafayette, IN 47907-2099, USA.

40% of the millennials and 35% of the other generations said they spent more on gasoline than 1 year ago; 29% of the millennials and 26% of the other generations said they spent more on automotive expenses (not including gas) than 1 year ago. Nelson[7] commented: "The millennials or Gen Y – will spend more than $200 billion annually starting in 2017, and $10 trillion in their lifetimes." Considering this trend of spending power, automotive companies wish to understand the characteristics of millennials and create cars (and the sounds they make) that are attractive to the next-generation car owners.

Compared to the Baby Boomers (born in the 1940s to 1950s) and Generation X (born in the 1960s to 1970s), millennials are very concerned about what others think of them[8], and have a higher level of self-esteem and self-expectations for the future[9]. Millennials are more fascinated by new technologies and adapt faster to computer and internet services[10,11]. Millennials enjoy interactive full motion multimedia, colorful images, and audio, and they prefer personalization and customization[11]. They are also more risk prone than members of Generation X according to Reisenwitz and Iyer[12]. All of these observations make it necessary to explore how we can integrate millennials' musical/sound preferences into the design of next-generation product sounds. Thus, in this research, we are primarily concerned with: (i) how millennials would choose between traditional sounds and those modified, innovative sounds in the given contexts and (ii) how millennials would evaluate these sounds.

Sound evaluation is a complex process and influenced by information from multiple sensory channels[13]. Many researchers have found that visual information can influence people's auditory perception. For example, Viollon et al.[14] studied the influence of visual settings on sound ratings, and found that, usually, the more urban the visual setting, the more negative the sound ratings. This result is also dependent on the type of sounds being considered. Menzel et al.[15] found that exposure to some colors can increase or decrease loudness judgements. Colors like red or pink seem to cause an increase in loudness, while grey or pale green was observed to decrease loudness. Yoshida et al.[16] presented luxury and sporty vehicle images to German and Japanese frequent drivers, while asking them to listen to the acceleration sounds. They found that German drivers tended to rate loudness lower and luxury higher when the sounds were presented with images of luxury cars rather than with images of sporty vehicles. Sportiness was rated higher when an image of a sporty vehicle was presented. From these studies, it is clear that the perception of an event in one modality is influenced by information presented in another modality[17]. Therefore, in this research, we were also interested in: (iii) how different visual cues of the given contexts (pictures or texts) would impact the millennial subjects' sound perception?

To answer these questions, an auditory perceptional test (hereinafter referred to as subjective test) was conducted to determine sound preferences when people were presented with the current sound, a very different sound and something in between the two. Intentional sounds (e.g., turn signal) and consequential sounds (e.g., car door closing) were considered in four car contexts and two other contexts. We also examined the influence of pictorial and textual information on those sound preferences and evaluations. This research may provide insights on how to integrate the generational differences of end-users into the design of future product sounds. In the remainder of this paper, the subjective test is described in Sec. 2, the test results are presented in Sec. 3, and the discussion and conclusions are provided in Sec. 4 and 5, respectively.

2 DESCRIPTION OF TEST

In this section, the creation of the test sounds, the overall structure of the test, the recruited human subjects and the test procedures are described.

2.1 Sound Stimuli

In this test, we developed a set of sound samples for the next-generation cars and other products. The basic idea was to create sounds with different "distances" from the current or traditional product sounds as shown in Fig. 1. Participants were presented with these three sounds and were asked to compare and evaluate them in the given context. Table 1 is a description of the 18 sounds created for the test, in which Contexts 1, 2, 5 and 6 are directly related to automobiles. Within each of the six contexts, three sounds were created: a "traditional" sound, a "very different" sound and a sound "between" them. Traditional sounds are recordings provided by automotive companies or other sources. Very different sounds are created by modifying the traditional sounds through use of signal processing techniques, including changing the frequency

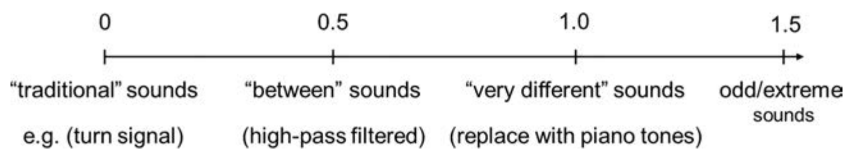

Fig. 1—A schematic diagram of the created sound samples with different distances from the current or traditional product sounds.

Table 1—A description of the sounds created for the test.

	Context	Sound	Sound description
1	Car door closing	1_1	"traditional"
		1_2	"between": add an ending beep
		1_3	"very different": replace with a musical tone
2	Turn signal	2_1	"traditional"
		2_2	"between": traditional sound high-pass filtered
		2_3	"very different": replace with piano tones
3	Put phone down	3_1	"traditional"
		3_2	"between": high pass filtered
		3_3	"very different": replace with pizzicato tones
4	Camera click	4_1	"traditional"
		4_2	"between": reversed signal
		4_3	"very different": replace with simple sine waves
5	Car horn	5_1	"traditional"
		5_2	"between": use different instrument tones
		5_3	"very different": use different instrument tones & rhythms
6	Windshieldwipers	6_1	"traditional"
		6_2	"between": enveloped random noise
		6_3	"very different": replace with musical instrument tones

or damping of components obtained from a Prony analysis of the signal, or by using musical tones and simulating the rhythm of the original signal. We chose to keep the duration of the three types of sounds almost the same and to preserve the major structural elements of the sounds. The musical tones were picked by considering the music or film preferences of millennials[18,19]. This method of creating "traditional," "very different" and "between" sounds represents one approach to create new sounds. We hope this study will inspire more innovative thinking in designing new product sounds beyond the method presented here.

For example, Sound 1_1 is a traditional car door closing sound (from a mass-produced sedan). Sound 1_2 is created by adding an ending beep (swept sine waves from 600 to 50 Hz) to Sound 1_1. Sound 1_3 is created by replacing the frequency components in the signal with a decaying musical tone (E1 = 41.2 Hz). The Prony analysis was first performed on Sound 1_1 and the frequencies of damped sine waves were altered. Similarly, Sound 2_1 is a traditional car turn signal, whereas Sound 2_3 comprises two piano tones from Adele's "Set Fire to the Rain" to imitate the "tick-tock" rhythm of the traditional turn signal. Sound 2_2 is a high-pass filtered version of Sound 2_1. Pilot tests showed that the "between" sounds (e.g., 1_2, 2_2, etc.) were considered to be close to the traditional sounds (e.g., 1_1, 2_1, etc.), while the "very different" sounds (e.g., 1_3, 2_3, etc.) were considered to be further away from the traditional sounds. The time histories and amplitude density spectrum of all 18 sounds are provided in Appendix. All sounds within a context were normalized to have, approximately, the same level.

2.2 Overall Structure of the Test

The subjective test consisted of five parts, as illustrated in Fig. 2. In the first part, participants just listened to the 18 sounds without rating them. In Parts 2 and 3 subjects compared, rated and described sounds. In both of these parts, they listened to sounds presented in six different contexts (three sounds were used for each context). Half the subjects did Part 2 (where images of the context were presented with no descriptive text) followed by Part 3 (where context was described in words with no images) and half the subjects did Part 3 followed by Part 2. Details of these parts are given below.

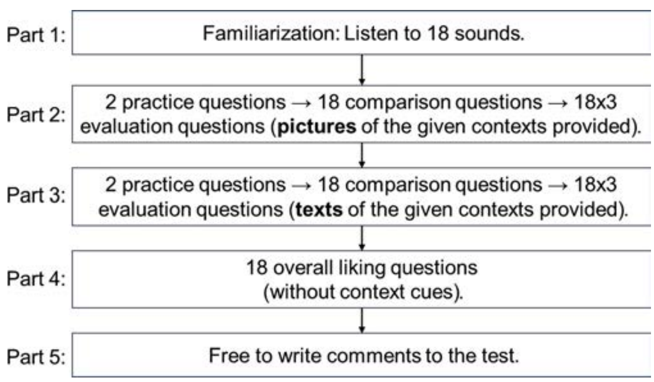

Fig. 2—The overall structure of the test.

To avoid the impact of colors, monochromatic pictures were used in Part 2. In Part 4, subjects rated how much they liked the 18 sounds presented without contextual cues. At the end of the test, subjects were asked to write comments on the test.

In Part 2, participants first did two practice evaluations which consisted of choosing one of two sounds (comparison question) and then doing an evaluation of a sound. These two practice evaluations contained sounds and contexts that were not used after the practice session. Subjects then proceeded to do 18 comparisons, 3 per context (both AB and BA presentations were used to account for ordering effects, but individual subjects were only presented with one of the pairs). The 18 context-sound pair groupings were presented in a different random order for each subject. Subjects were asked to respond to: "If you had a choice, which sound would you prefer to hear in the following situation?". Figure 3(a) is an illustration of

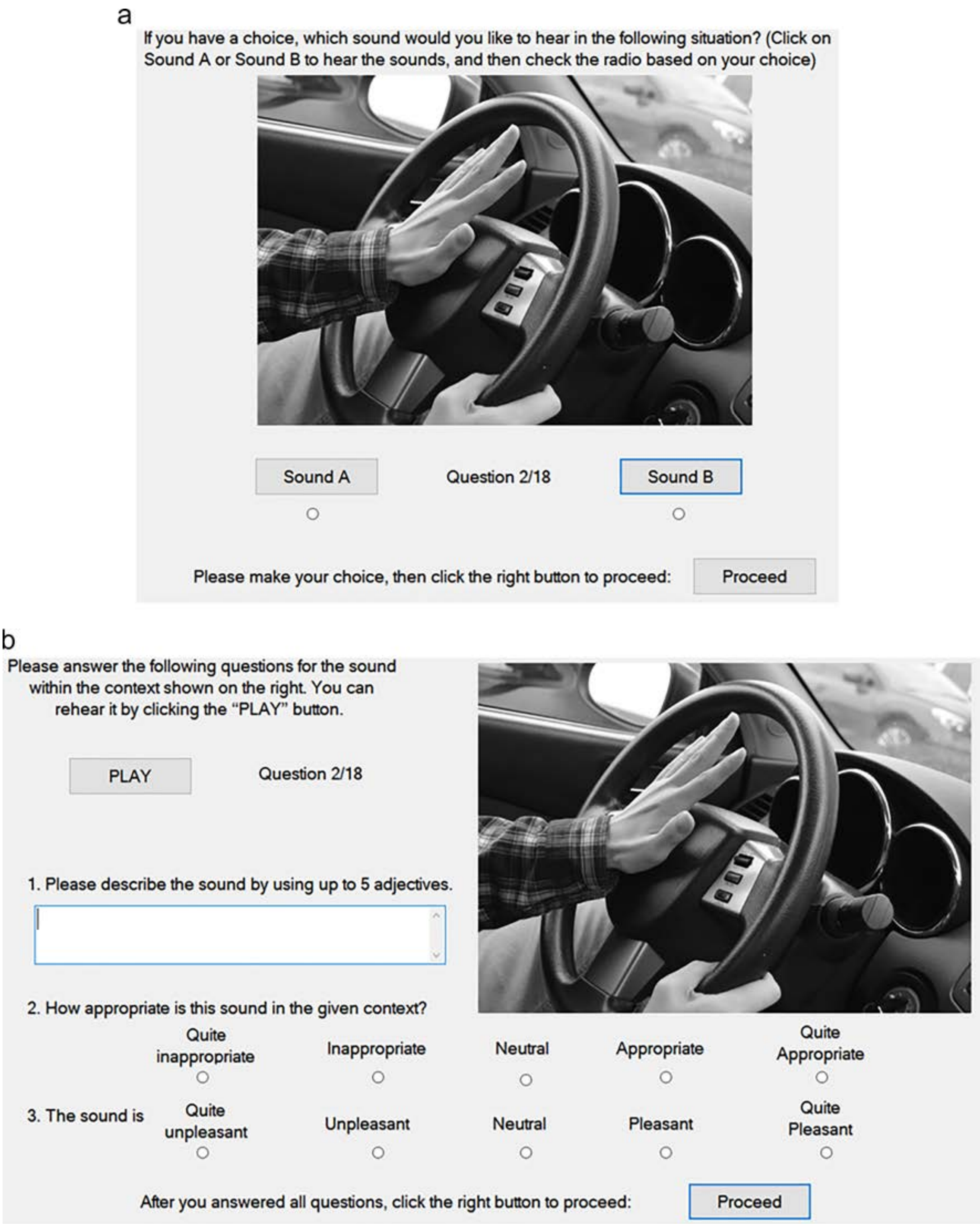

Fig. 3—The screenshots of the test program: (a) Comparison question; (b) Evaluation questions (developed with Microsoft Visual C++ 2010).

the interface that the subjects would see. On completion of the comparisons, they answered a set of three evaluation questions for each of the 18 sounds. Again, contextual information was provided as shown in Fig. 3(b). Subjects were asked to: (a) describe the sound by using up to 5 adjectives, (b) indicate how appropriate is this sound in the given context, and (c) indicate how pleasant they found the sound? The appropriateness question provides five options to the participants, varying from “quite inappropriate,” “neutral” to “quite appropriate”. Similarly, participants could choose from five options for the pleasantness question, varying from “quite unpleasant,” “neutral” to “quite pleasant”.

Part 3 is basically the same as Part 2, except that text is used to describe the contexts instead of providing pictures. For example, instead of seeing a picture of the “car horn” context, participants saw a sentence: “You are sounding the horn of your car,” when answering the comparison and evaluation questions in this context. Therefore, participants did each comparison and sound evaluation twice: once with pictorial cues of contexts and once with textual cues.

In Part 4, participants heard all sounds again and were asked how they liked them. Four options were provided, “I really dislike it,” “I dislike it,” “I like it” and “I really like it.” They clicked on a button to input their answers. Again, all sounds were played in a different random order for each subject and without any pictorial or textual contextual cues. As noted above, Part 2 and Part 3 were switched for half of the participants; thus, we have 2 main groups and 4 subgroups of participants. In group 1, participants saw pictorial cues of contexts first (subgroup “P1: Pic_First”) and then textual cues (subgroup “T2: Text_Second”), whereas, in group 2, participants saw textual cues of contexts first (subgroup “T1: Text_First”) and then pictorial cues (subgroup “P2: Pic_Second”).

In Fig. 4, shown is a schematic diagram of the range of the appropriateness ratings and overall liking ratings. Five options were given in the appropriateness and pleasantness questions, so these two ratings range from -2 to 2. We intentionally provide four options for the overall liking questions, to push participants to express their general opinions on the sounds. To keep a consistent difference between two consecutive options, the overall liking ratings range from -1.5 to 1.5.

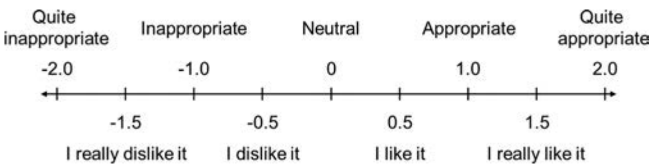

Fig. 4—A schematic diagram of the range of the appropriateness ratings and overall liking ratings.

2.3 Test Procedures

Forty university students and staff participated in the test voluntarily. Thirty-five participants were younger than 36 years old, i.e., they belong to the millennial generation. These millennials include 17 females and 18 males ranging in age from 18 to 32 years. Their median age was 22 years (mean = 23.3 years, SD = 3.9 years). Subjects' awareness of sound quality and noise control ranged from nothing to moderate and none had taken acoustics or noise control courses. Six subjects had worked in noisy industries (airport, engine testing or concerts), or regularly used firearms; five subjects had previously been involved in sound quality tests or vibration/noise control studies; eleven subjects had studied music and/or had been involved in musical events or activities.

The test was conducted in a double walled sound booth in the Ray W. Herrick Laboratories. The playback system consisted of a LynxOne sound card, Tucker-Davis HB7 amplifier and Etymotic Research ER-2 tube earphones. Prior to subject arrival, the left and right channels were calibrated using 85 dB and 1 kHz calibration tones with the same calibration factor as the signals used in the test. The maximum A-weighted sound pressure level of each signal for each channel was recorded to check the calibration of the signals and to ensure that the sounds were presented to the subjects at a safe listening level.

The subject first read and signed a consent form and filled out a questionnaire on his/her background. Then, the subject was given a hearing screening, to ensure that the subject's hearing thresholds were 20 dB or lower in all octave bands from 125 Hz to 8 kHz. The subject was then given the test instructions and a set of earphones and instructed on how to insert the earphones. The test then started. After completing the test, the subject was given a second hearing screening and compensated \$10 for participating in the test. The whole process took about 1 hour on average.

3 RESULTS

Thirty-five of the 40 participants were millennials, and in this paper, we only present the test results from the millennial participants. In this section, results of the paired comparisons and sounds evaluations (Appropriateness, Pleasantness and Overall Liking) are presented. The semantic analysis of the sound descriptions and the relationships between sound preferences, evaluations and descriptions are provided. The influence of the visual contextual cues on the sound evaluations is also given.

3.1 Paired Comparisons

Within each context, participants performed three paired comparisons in order to select the sound that better matched the context. The Bradley–Terry Logit (BTL) model was used to determine the relative strength of preference for the paired comparisons[20]. This model states that the probability that option "i" is chosen over option "j" ($P_{(i>j)}$) can be modeled by Eqn. (1), where p_i is a positive real-valued score assigned to option "i." p_i may represent the strength of preference of the option "i."

$$P_{(i>j)} = \frac{p_i}{p_i + p_j} = \frac{e^{\beta_i}}{e^{\beta_i} + e^{\beta_j}} = \frac{1}{1 + e^{(\beta_j - \beta_i)}}. \qquad (1)$$

If the probabilities can be estimated, then the strength of each item can be estimated by using Eqn. (1). Hunter's MM algorithm[21] was used to fit the Bradley–Terry Logit model and calculate the BTL preference values (β_i, β_j). The estimated BTL preference values for the 18 sounds when combining the comparisons across all 4 subgroups are shown in Table 2. The additional constraint used to determine the BTL preference values is that the BTL preference value of the traditional sound, in each context, is set to 0 ($p_1 = 1$); all other BTL preference values are either greater than or less than 0. A sound with a larger BTL preference value is more likely to be chosen as being more preferable in paired comparisons within the given context. As shown in Table 2, in Contexts 1, 2, 4 and 5, participants were less likely to choose modified and novel sounds, whereas, in Context 6, they were more likely to choose them. In Context 3, participants seem to prefer the "between" sounds more than the "traditional" and "very different" sounds. In all of the six contexts, participants showed higher preferences for Sound 2 ("between") than for Sound 3 ("very different"), and in Contexts 2 and 4, this preference is stronger.

3.2 Sound Evaluations

Participants' evaluations (appropriateness, pleasantness and overall liking) of the sounds when combining the data across all 4 subgroups and the relationships between the evaluations are presented in this subsection.

Table 2—The BTL preference values of the 18 sounds in 6 contexts, estimated from the subjects' selections of sounds based on which they would prefer to hear in a context.

No.	Context	Sound 1 "traditional"	Sound 2 "between"	Sound 3 "very different"
1	Car door closing	0.00	−1.08	−1.27
2	Turn signal	0.00	−0.22	−1.50
3	Put phone down	0.00	0.76	−1.19
4	Camera click	0.00	−0.97	−2.62
5	Car horn	0.00	−0.98	−1.81
6	Windshield wipers	0.00	1.21	1.15

3.2.1 Appropriateness ratings

Figure 5 shows the average of the appropriateness ratings (Appropriateness) of each sound for each of the 6 contexts. The error bars show the standard error of the average estimates. The appropriateness rating values range from −2 (quite inappropriate) to +2 (quite appropriate) and 0 means neutral.

Generally, participants considered "traditional" sounds (#_1) to be more appropriate in Contexts 1, 2, 4 and 5, and the "very different" sounds (#_3) to be inappropriate in all contexts. The "between" sounds (#_2) were generally rated to be more appropriate than the "very different" sounds and less appropriate than the "traditional" sounds, except in Contexts 3 and 6 where the "between" sounds were considered to be more appropriate than others.

3.2.2 Pleasantness ratings

In Fig. 6, shown is the average of the pleasantness ratings (Pleasantness) along with the standard error. Again, results combining all 4 subgroups within 6 contexts are shown. The pleasantness ratings range from −2 (quite unpleasant) to +2 (quite pleasant) and 0 means neutral. In general, people gave positive pleasantness ratings to the sounds in Contexts 1, 2 and 4, but negative ratings in

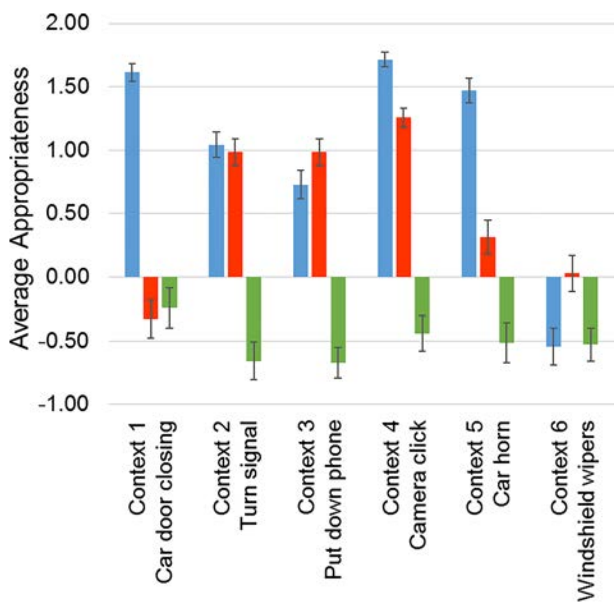

Fig. 5—The average of the appropriateness ratings of each sound. Bars indicate the ± standard error. Results for original (Sound #_1), modified (Sound #_2) and novel sounds (Sound #_3) are in blue, red and green, respectively.

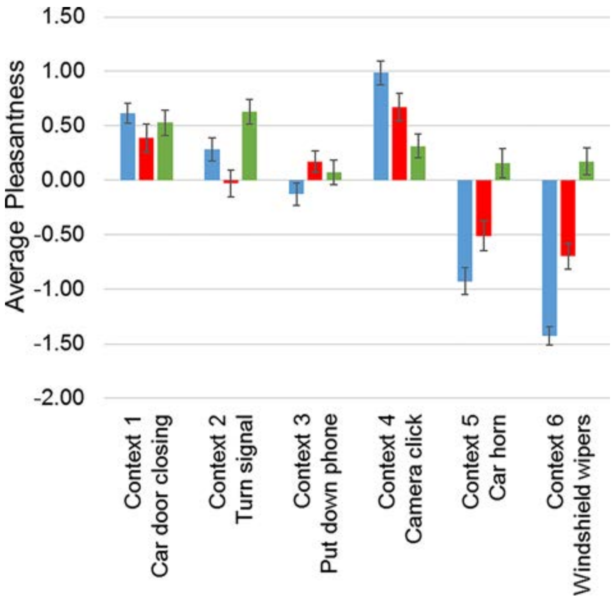

Fig. 6—The average of the pleasantness ratings of each sound. Bars indicate the ± standard error. Results for original (Sound #_1), modified (Sound #_2) and novel sounds (Sound #_3) are in blue, red and green, respectively.

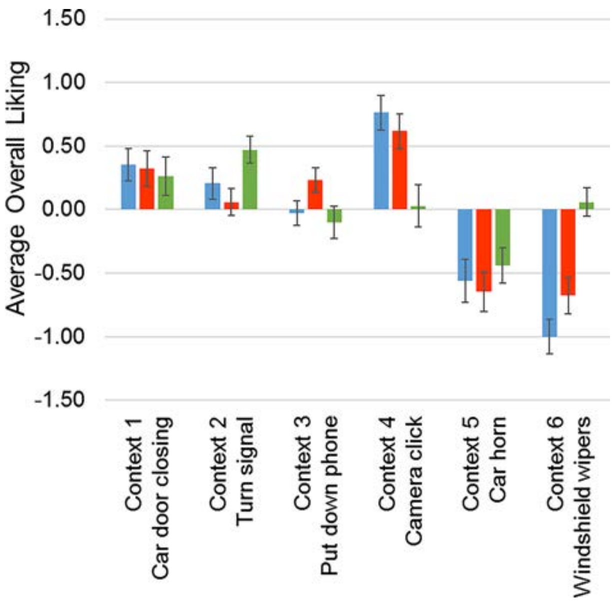

Fig. 7—The average of the overall liking ratings of each sound. Bars indicate the ± standard error. Results for original (Sound #_1), modified (Sound #_2) and novel sounds (Sound #_3) are in blue, red and green, respectively.

Contexts 5 and 6. Notice that participants gave the Sound 2_3 the highest pleasantness rating (this sound consists of two piano tones); however, it received a very low rating on its appropriateness (see Fig. 5).

3.2.3 Overall Liking ratings

In Fig. 7, the average overall liking ratings are shown along with standard error bars for all the sounds presented within 6 contexts. The overall liking evaluations range from −1.5 (I really dislike it) to +1.5 (I really like it). The overall liking results are consistent with the pleasantness results presented above. Again, Sound 2_3 (the musical turn signal) received the highest Overall Liking rating in the group of all "very different" sounds (green bars), although its BTL preference value and average appropriateness rating are very low (see Table 2 and Fig. 5).

3.2.4 Correlations between sound evaluations and preferences

In Fig. 8, the average Appropriateness and BTL preference values of each sound are plotted against each other. Note that, here, the BTL preference values of all "traditional" sounds (i.e., Sound #_1, # is the context number) were set to their appropriateness values and the BTL

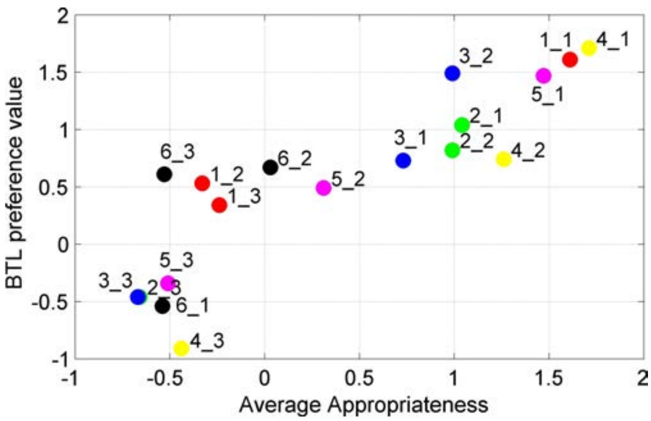

Fig. 8—Relationship between BTL preference value and average Appropriateness of each sound (R^2 = 0.551, excluding the signals with their BTL preference values set to their corresponding appropriateness values). Red, green, blue, yellow, pink and black represent sounds in Contexts 1 (car door closing), 2 (turn signal), 3 (put phone down), 4 (camera click), 5 (car horn) and 6 (windshield wipers), respectively. Note here that the BTL preference values for #_1 sounds are set to the average appropriateness ratings.

preference values of other sounds were adjusted accordingly. This means that R^2 values will likely be higher if subjects had rated preference within a context on a rating scale rather than in a paired comparison test, because the traditional sounds here are made equal on the BTL preference and appropriateness scales. Statistics show that average Appropriateness and BTL preference values are significantly correlated ($R^2 = 0.551, p < 0.001$, excluding the signals with their BTL preference values set to their corresponding appropriateness values). It is interesting to observe that Sound 6_3 was deemed not appropriate but was preferred in this context. This may be because Sound 6_1 is an example of poorly functioning windshield wiper and the picture may also have given an impression of this being a "good" car, thus creating a mismatch between visual and acoustic information. Correlations between Pleasantness and BTL preference values and correlations between Liking and BTL preference values were not significant (see Fig. 9). It is concluded that participants were more likely to choose the sounds that they thought to be appropriate in the given contexts even when the sounds were not pleasant to them or they disliked them.

In Fig. 10, the average Pleasantness and average Overall Liking of each sound are plotted against each other. Statistics show that average Pleasantness and average Overall Liking are significantly correlated ($R^2 = 0.872$, $p < 0.001$). This result illustrates the strong connection of participants' evaluation in pleasantness and overall liking of sounds. Correlations between Appropriateness and Pleasantness values, and correlations between Appropriateness and Overall Liking values are insignificant. These results are similar to those in Fig. 9. From these results, it is concluded that appropriateness may not be necessarily related to the pleasantness or liking of a sound.

3.3 The Sound Descriptions

In the subjective test, participants' verbal descriptions of each sound were collected. As shown in Table 3, we classified these descriptions into nine categories. The first seven categories were each split into two subcategories, which represent opposites of one another. For example, the "loudness" category is split into "loud" and "soft." For each sound and each participant, we calculated how many times each participant's descriptions (words or groups of words) fell into any of the 16 subcategories. Then for each subcategory, a positive or negative number was assigned for one occurrence of that description. For example, 1 point was assigned if one's descriptions belong to the subcategory "loud," while -1 point was assigned if one's descriptions fall into the subcategory "soft." These numbers were then summed for each sound. In this way, a numerical value for each sound in that category was obtained. For example, a sound with a positive loudness value means that this sound received more descriptions of "loud" than "soft," i.e., participants generally considered this sound to be loud. Similarly, sounds with higher values in the "emotion," "duration," "spectral balance" and "novelty" categories were considered to be more pleasant, longer, sharper and more futuristic by participants, respectively. Table 4 is a summary of the calculated values for each sound and each category based on participants' verbal descriptions of the sounds.

In Fig. 11, the normalized category values for each sound based on Table 4 are plotted (seven important

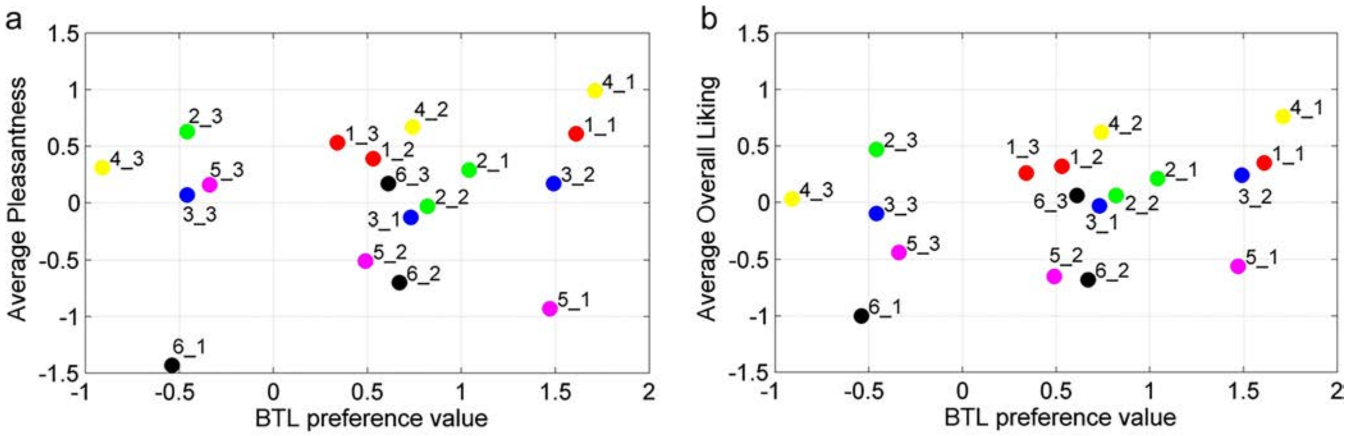

Fig. 9—Relationship between BTL preference value and average pleasantness rating and overall liking rating for each sound: (a) Pleasantness vs. BTL, R^2 =0.030; (b) Overall Liking vs. BTL, R^2 = 0.108. Red, green, blue, yellow, pink and black represent sounds in Contexts 1 (car door closing), 2 (turn signal), 3 (put phone down), 4 (camera click), 5 (car horn) and 6 (windshield wipers), respectively. Note here that the BTL preference values for #_1 sounds are set to the average appropriateness ratings.

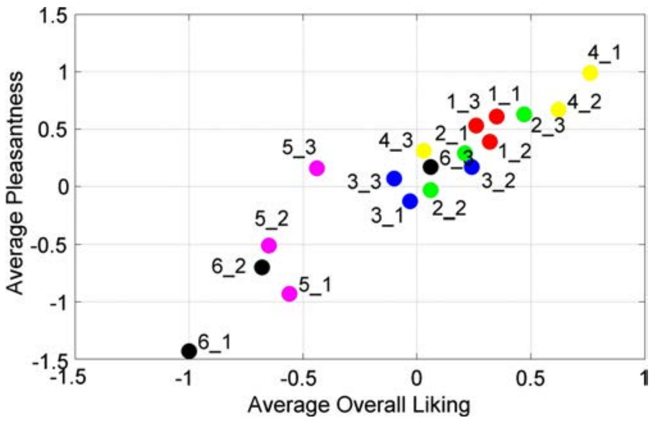

Fig. 10—Relationship between average Pleasantness and average Overall Liking of each sound, R^2 =0.872. Red, green, blue, yellow, pink and black represent sounds in Contexts 1 (car door closing), 2 (turn signal), 3 (put phone down), 4 (camera click), 5 (car horn) and 6 (windshield wipers), respectively.

pairs of categories are shown). Note that the musical category scores for Sound 2_3 and Sound 5_3 are higher than for other sounds because these innovative “very different” sounds were for the most part, intentionally musical. These sounds were deemed more pleasant (by word category score) than other sounds, but the word scores in the suitability category are low. Sound 6_1 was considered to be unpleasant and inappropriate, which received more descriptions in the categories of “sharp,” “loud” and “annoying” than the modified and novel sounds. This is consistent with the results of the Appropriateness and Pleasantness ratings discussed in Sec. 3.2.4.

In Table 5, the significant correlations of the category values derived from participants' sound descriptions (data in Table 4) and the results from the paired comparisons and the ratings tests are given. We observe that the “suitability” category values are positively correlated with the BTL preference values and appropriateness ratings. Sounds with higher “emotion” category values are, on-average, rated with higher Pleasantness and Overall Liking by participants. This indicates that participants' sound descriptions are consistent with their sound evaluations. Also, we found that the sounds with more descriptions in the category “novelty” tend to receive lower ratings in Appropriateness and lower BTL preference within context values. Participants did not like the sounds that were too loud and sharp, which is not unexpected given what is generally known about how these attributes affect annoyance (see, for example, Zwicker and Fastl's Psychoacoustic Annoyance model[22]).

$$\begin{aligned}\text{Appropriateness} = &-0.769\\ &-0.040\,\text{emotion}\\ &-0.028\,\text{duration}\\ &+0.080\,\text{suitability},\end{aligned} \quad (2)$$

$$\begin{aligned}\text{Pleasantness} = &\ 0.522 - 0.007\,\text{loudness}\\ &+0.027\,\text{emotion}\\ &+0.018\,\text{suitability}\\ &-0.046\,\text{simplicity},\end{aligned} \quad (3)$$

Table 3—Categories and coding of the sound descriptions.

Category	Subcategory	Examples
Loudness	Loud (+1)	Loud, heavy, strong, hard, forceful
	Soft (−1)	Soft, light, quiet, calm, smooth, calming
Emotion	Pleasant (+1)	Pleasant, happy, pleasing, exciting, comforting
	Annoying (−1)	Annoying, irritating, unpleasant, unhappy, noisy
Duration	Long (+1)	Echo, long, repetitive, slow, lingering
	Quick (−1)	Quick, short, fast, brief, rapid, swift
Spectrum balance	Sharp (+1)	Sharp, high-pitched, harsh, rough, scratching
	Dull (−1)	Dull, deep, low, muffled
Novelty	Futuristic (+1)	Modern, artificial, new, unusual, alien, techo
	Normal (−1)	Normal, general, ok, medium, neutral, common
Suitability	Appropriate (+1)	Appropriate, good, nice, satisfying, accurate
	Inappropriate (−1)	Inappropriate, weird, odd, wrong, ridiculous
Simplicity	Simple (+1)	Simple, clean, concise, monotone, thin
	Changing (−1)	Changing, rich, sophisticated,
	Grab-attention	Grab-attention, noticeable, alerting, warning, distinct
	Musical	Musical, toned, melodic, rhythmic, orchestral

Table 4—Category values for each sound based on sound descriptions, where spec. bal. is spectrum balance, suit. is suitability, and grab. is grab-attention.

Context	Sound	Loudness	Emotion	Duration	Spec. bal.	Novelty	Suit.	Simplicity	Grab.	Musical
Car door closing	1_1	19	−5	−26	−8	−24	21	4	2	0
	1_2	−3	3	−25	−6	38	−5	5	2	1
	1_3	−29	3	−38	−13	28	−3	8	2	1
Turn signal	2_1	−13	−12	−2	−9	−19	15	7	10	2
	2_2	−8	−16	−13	25	−10	8	8	6	2
	2_3	−8	6	−1	5	−8	7	7	3	45
Put phone down	3_1	27	−12	−17	−4	−13	10	5	2	2
	3_2	−29	−4	−35	4	−18	2	9	2	2
	3_3	−16	−11	−27	−4	14	−11	1	4	3
Camera click	4_1	−11	11	−15	7	−22	29	9	8	2
	4_2	−19	4	−21	18	−11	19	6	0	2
	4_3	−13	−2	−17	11	38	2	3	4	1
Car horn	5_1	49	−43	−12	16	−14	0	3	10	8
	5_2	15	−26	−3	30	8	2	5	10	23
	5_3	−1	6	2	2	5	8	12	5	29
Windshield wipers	6_1	34	−45	−2	27	−2	−17	2	4	0
	6_2	28	−26	5	0	−6	−4	2	1	1
	6_3	−8	−9	−1	−12	22	−7	4	1	20

$$\text{Overall Liking} = -0.042 + 0.016\,\text{emotion} - 0.010\,\text{duration} + 0.014\,\text{suitability}. \quad (4)$$

Equations (2) to (4) are the estimated regression models to predict Appropriateness, Pleasantness and Overall Liking using the categories values given in Table 4.

The R^2_{adjusted} of these models is 0.896, 0.952 and 0.770, respectively, and all coefficients in the regression models are significant at the 95% confidence level. Since this is an exploratory study, it is of importance to note that

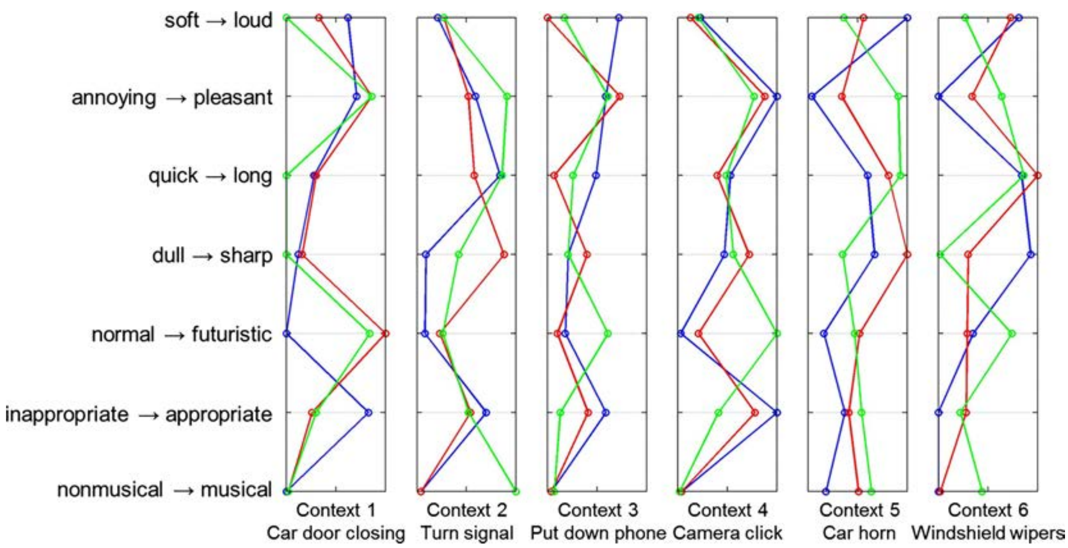

Fig. 11—Normalized category values for each sound based on sound descriptions. Results for original (Sound #_1), modified (Sound #_2) and novel sounds (Sound #_3) are in blue, red and green, respectively. The closer a data point is to the right hand side, the more descriptions this sound received in the corresponding category (e.g., "loud"), and vice versa.

Table 5—Significant correlations between scores derived from sound descriptions and sound evaluations. BTL preference values are a measure of preference within a particular context.

Correlations		Corr.	*p*-Value
Positive	BTL vs. suitability	0.544	0.020
	Appropriateness vs. suitability	0.734	0.001
	Pleasantness vs. emotion	0.945	<0.001
	Liking vs. emotion	0.825	<0.001
Negative	BTL vs. novelty	−0.616	0.006
	Appropriateness vs. novelty	−0.742	<0.001
	Pleasantness vs. loudness	−0.739	<0.001
	Pleasantness vs. spectrum balance	−0.475	0.046
	Liking vs. loudness	−0.677	0.002
	Liking vs. duration	−0.500	0.035

these regression models need to be validated by collecting participants' response to additional proposed sounds in future. These models provide some insight into creation of new product sounds. For example, suitability and emotion both play important roles in the Appropriateness, Pleasantness and Overall Liking ratings in our test.

3.4 The Impact of Pictorial versus Textual Context Cues on Sound Evaluations

In the subjective test, the participants were divided into 2 main groups and 4 subgroups. Group 1 saw the pictorial contextual cues first when answering the sound comparison questions and evaluation questions, and then saw the textual cues when answering these questions, the second time. Group 2 saw textual cues first and then pictorial cues. Thus, as mentioned earlier, we have four subgroups of participants for answering the sound comparison and evaluation questions: “P1: Pic_First,” “T2: Text_Second,” “T1: Text_First” and “P2: Pic_ Second.” One-way ANOVA and a Tukey test with a 95% confidence level were used to detect the differences of these sound evaluations between different subgroups. In most cases, we did not find statistically significant differences and significant group rating differences are given in Table 6. Under each subgroup, the mean rating value and the corresponding standard error are given. Subgroups that yield average ratings that are significantly different from the average ratings for the other subgroups are denoted by a different number of asterisks (*), i.e., the average rating in a subgroup denoted by one asterisk (*) is significantly different from the rating in a subgroup denoted by two asterisks (**). The average ratings in two subgroups denoted by one asterisk (*) or without asterisks are not significantly different from each other. For example, the average Appropriateness ratings of sound 5_1 is 1.82 with a standard error of 0.10 in subgroup P1 “Pic_First,” while in subgroup P2 “Pic_Second,” it is 1.06 with a standard error of 0.22. This difference is statistically significant ($F(3, 71) = 4.49$, $p = 0.006$), which suggests that P1 participants who saw the pictorial cues before textual

Table 6—Summary of the significant group differences in sound comparisons and evaluations. For each sound listed, the average ratings for each subgroup (N = 20) and the standard error are reported. Sounds 1_ #, 3_# and 5_# belong to the Contexts 1 (car door closing), 3 (put phone down) and 5 (car horn) respectively. Subgroup ratings with a different number of asterisks (*) are significantly different from each other in the Tukey test, i.e., the average rating in a subgroup denoted by one asterisk (*) is significantly different from the rating in a subgroup denoted by two asterisks (**).

	Sound	P1	T2	T1	P2	*F*-value	*p*-Value
Selected times in	5_1	1.77 (0.14)	1.82 (0.13)**	1.61 (0.16)	1.17 (0.20)*	3.40	0.023
paired comparisons	5_3	0.30 (0.11)	0.12 (0.08)**	0.50 (0.19)	0.83 (0.22)*	3.61	0.018
Appropriateness	5_1	1.82 (0.10)**	1.77 (0.11)	1.28 (0.23)	1.06 (0.22)*	4.49	0.006
Pleasantness	1_1	0.77 (0.18)	1.00 (0.15)**	0.44 (0.19)*	0.28 (0.19)*	3.24	0.027
	5_2	−0.94 (0.19)**	−0.77 (0.26)	−0.06 (0.29)*	−0.44 (0.29)	2.73	0.051

cues tended to consider Sound 5_1 to be more appropriate in the context than the P2 participants who saw the pictorial cues second.

The Overall Liking questions were answered by participants without seeing any contextual cues while the sounds were being played, although they had seen context cues in earlier parts of the test. One-way ANOVA and Tukey test were used to detect the differences of Overall Liking ratings between Group 1 and Group 2 and no significant differences were found.

In Fig. 12, shown are the relationships between Pleasantness and Overall Liking ratings for the four subgroups. The R^2 for the relationships in subgroups "P1: Pic_First," "T2: Text_Second," "T1: Text_First" and "P2: Pic_Second" is 0.941, 0.876, 0.651 and 0.731, respectively. In Fig. 13, shown are the same relationships for the two main groups, Group 1 (P1 and T2) and Group 2 (T1 and P2). The R^2 for the relationships in Group 1 and Group 2 are 0.929 and 0.714, respectively. The positive correlations between these two ratings for Group 1 are much higher than for Group 2 when all Group 1 results are combined and all Group 2 results are combined. This reveals that the order of showing picture and text contextual cues could impact the consistency of participants' evaluations of sounds.

4 DISCUSSION

In most cases, millennial participants tended to choose more traditional sounds in the given contexts. Even when choosing the "between" sounds and the "very different" sounds, they still preferred the sounds closer to the traditional ones. One exception is the traditional Sound 6_1 in Context 6 (windshield wipers). It appears that the high frequency part of the original sound made participants feel uncomfortable, and probably made them think that the mechanical parts of the wipers may not work very well. Some participants commented on this sound as a signal of a pair of "broken wipers," though the picture of a pair of well-performing wipers was shown in the test. In contrast, the modified sounds (Sound 6_2 and Sound 6_3)

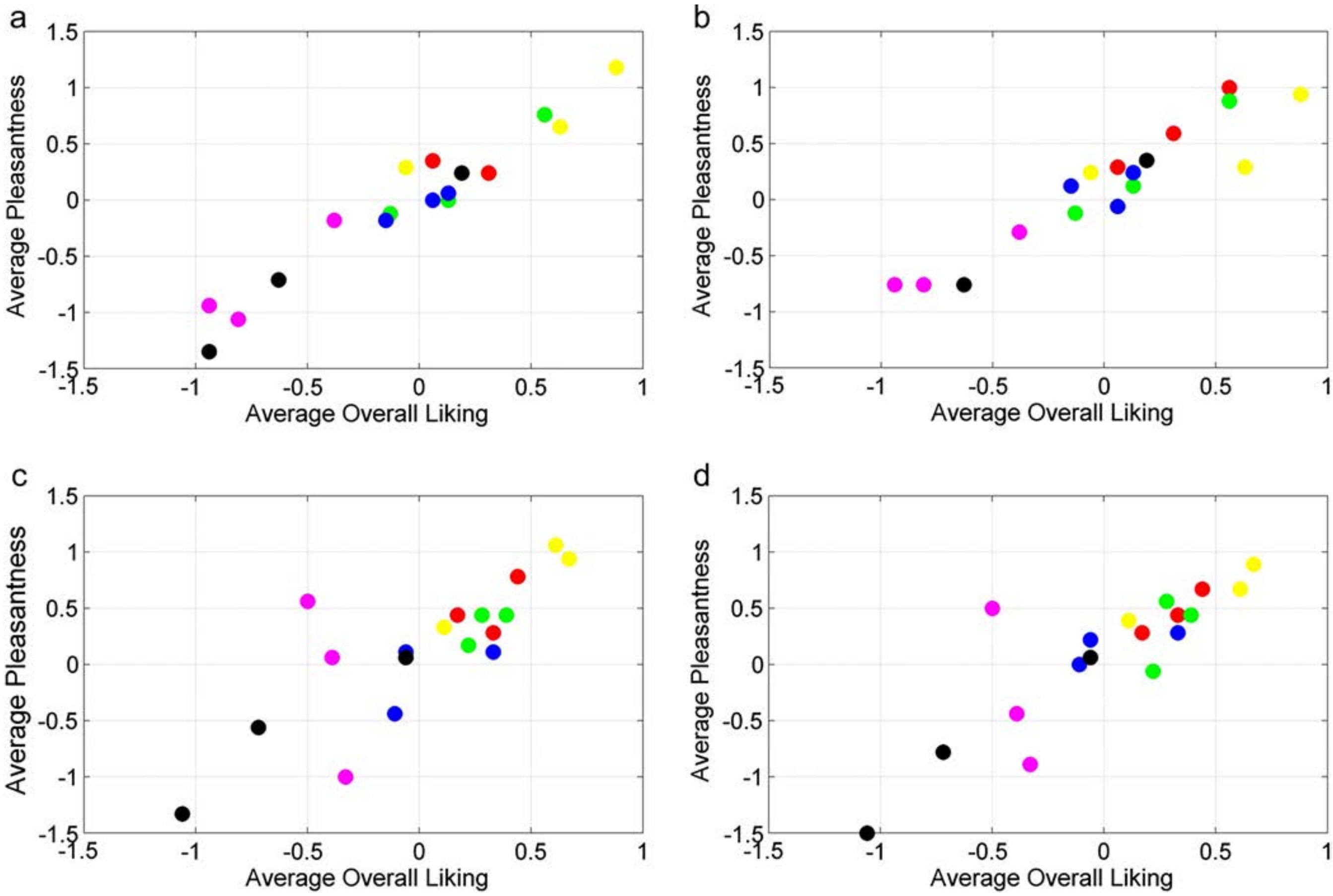

Fig. 12—The relationships between pleasantness and overall liking ratings in 4 subgroups: (a) Subgroup "P1: Pic_First," R^2 = 0.941; (b) Subgroup "T2: Text_Second," R^2 = 0.876; (c) Subgroup "T1: Text_First," R^2 = 0.651; (d) Subgroup "P2: Pic_Second," R^2 = 0.731. Red, green, blue, yellow, pink and black represent sounds in Context 1 (car door closing), 2 (turn signal), 3 (put phone down), 4 (camera click), 5 (car horn) and 6 (windshield wipers), respectively.

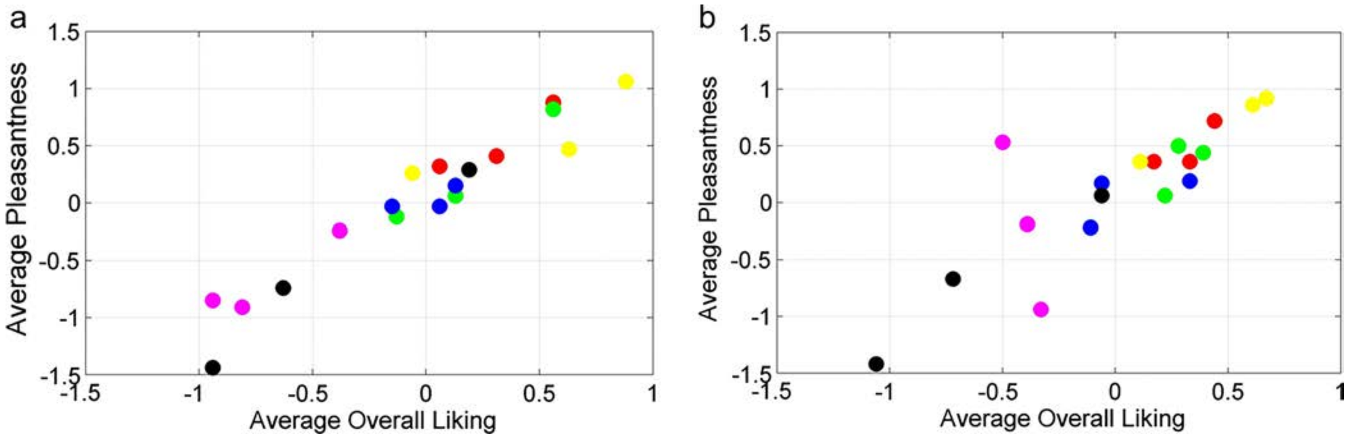

Fig. 13—The relationships between pleasantness and overall liking ratings in 2 main groups: (a) Group 1, R^2 = 0.929; (b) Group 2, R^2 = 0.714. Red, green, blue, yellow, pink and black represent sounds in Context 1 (car door closing), 2 (turn signal), 3 (put phone down), 4 (camera click), 5 (car horn) and 6 (windshield wipers), respectively.

did not have very much high frequency content due to the low-pass filtering of the sound and the frequencies chosen in the creation of the innovative sound. Participants described these sounds as gentler and smoother, which could explain why participants chose them instead of the traditional sound.

Sound evaluations provide more insight into how participants perceive and judge these sounds. Generally, participants considered "traditional" sounds to be more appropriate in Contexts 1, 2, 4 and 5, and the "very different" sounds to be inappropriate in all 6 contexts. This result is consistent with their choices in the paired comparisons and their sound descriptions. The average Appropriateness and the BTL preference values of each sound are positively correlated (see Fig. 8). Additionally, we did not find significant correlation between Appropriateness and Pleasantness values, between Pleasantness and BTL preference values, nor between Overall Liking and BTL preference values.

Participants were more likely to choose the sounds that they thought better matched the given contexts, even though these sounds were not pleasant to them. For example, in Context 5 (car horn), some participants stated "this sound doesn't have to be too pleasant because the horn sound is mainly to grab others' attention." Another example is the Sound 2_3 which consists of two piano tones from a popular song clip. This sound had the highest Overall Liking rating; however, its Appropriateness and BTL preference values are quite low. From this discussion, we may conclude that the priority in the design of product sounds is to convey the correct signal/function/meaning of the context to listeners. If the sound fails to convey such information or the sound conveys incorrect information (e.g., Sound 6_1 with the "broken wipers"), people will likely not find them appropriate in that context.

Participants' verbal descriptions of the sounds are consistent with their evaluation ratings and also provide subtler feelings and judgments of the sounds. We observe that participants considered the simple and appropriate sounds to be more pleasant to them, while the loud and sharp sounds had the reverse effect. The regression models to predict the average Appropriateness and Pleasantness ratings using the frequency of occurrence of the sound descriptors within a category provide sound designers with insights that may help them modify and improve the created sounds. More importantly, the 16 categories of the sound descriptions can be used as the basis for semantic differential test scales in the future, or can be used to help gather more consistent feedback from customers who dislike certain car sounds.

One particular goal of this study is to investigate how different types of contextual cues (picture or text) and their presenting order would influence participants' perception and evaluation of the sounds. However, for most cases in this test, this influence is insignificant, for only a very few sounds is this significant. One explanation could be that, in this test, the more innovative sounds attract much more attention than the contextual cues. Participants may have focused on the quality and attributes of the sounds and did not pay much attention to the differences in contextual cues. But still, we found for a few sounds, on average, participants seeing different types of contextual cues had significantly different ratings of the sound's Appropriateness or Pleasantness. Thus, visual contextual cues sometimes may play an important role in people's sound perceptions, which is worthy of consideration when sound designers present different sounds to customers. The group that saw the picture cues first and text cues second produced more consistent results across parts of the test.

5 CONCLUSIONS

In summary, millennials are more likely to accept the traditional sounds in the given contexts. Modified or innovative sounds might be interesting or even more pleasant to them; however, when they found these sounds to be inappropriate in the contexts, they did not prefer them. The exception to this finding was when the original sound represented a poor quality version of the product and the picture shown was of a more luxurious car; in this case, people preferred the in-between sound and the innovative sound was also preferred over the original sound. Participants' sound preferences were found to correlate with their Appropriateness ratings positively, and their Pleasantness ratings correlate with their Overall Liking positively. Participants' verbal descriptions of the sounds are aligned with their sound evaluations as well as provide information for construction of semantic scales in future research. The pictorial and textual cues of context and their presentation order can impact how people perceive sounds in some cases. In general, subjects' responses were more consistent across all the tests when shown pictorial cues first; this emphasizes the need to be careful in how context cues are presented in subjective tests. The results of this study are useful for designers of future product sounds who wish to consider end users' sound preferences in the product sound design.

A limitation of this study is the lack of consideration of how millennials might learn or adapt to new sounds. If they hear the "very different" sounds for a certain amount of time, we do not know if they would still mostly hold negative opinions on the acceptability of those sounds in particular contexts. Future research opportunities include collecting more sound samples in a particular context (e.g., car door closing) and investigating how the signal components (spectral balance, loudness, etc.), visual representation of contexts (pictures, videos, virtual reality simulation, and physical contact), and participants' background (age, gender, driving experience, etc.) influence people's preferences of these sounds. Contextual information should be considered carefully as people may not desire to hear too pleasant sounds under certain circumstances (e.g., car horn). While results from the 36 subjects who were millennials were presented in this paper, it would be interesting to test a larger group of older subjects to compare their responses to the millennial group.

6 ACKNOWLEDGMENTS

Sincere gratitude is extended to Daniel Carr, Jelena Paripovic and Weonchan Sung, who provided valuable technical support. We also would like to thank John Nalevanko, Haleh Ghalambor, Joshua Wheeler and Arun Mahapatra from Ford Motor Company for their great help in this project.

7 APPENDIX

In this appendix are plots of the time histories of the sounds and the associated amplitude density spectra

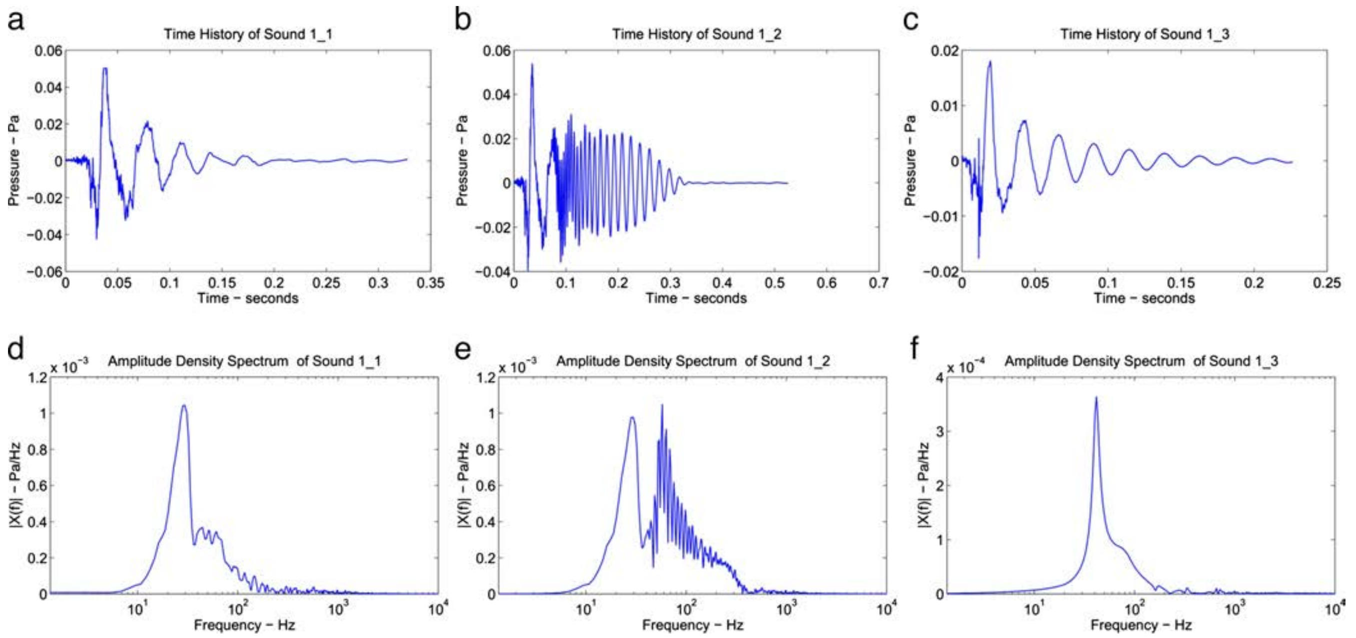

Fig. A1—The time histories and amplitude density spectrum of the sounds in Context 1 (car door closing). Subplots (a)–(c) are time histories of the "traditional," "between" and "very different" sounds, respectively. Subplots (d)–(f) are the associated amplitude density spectra.

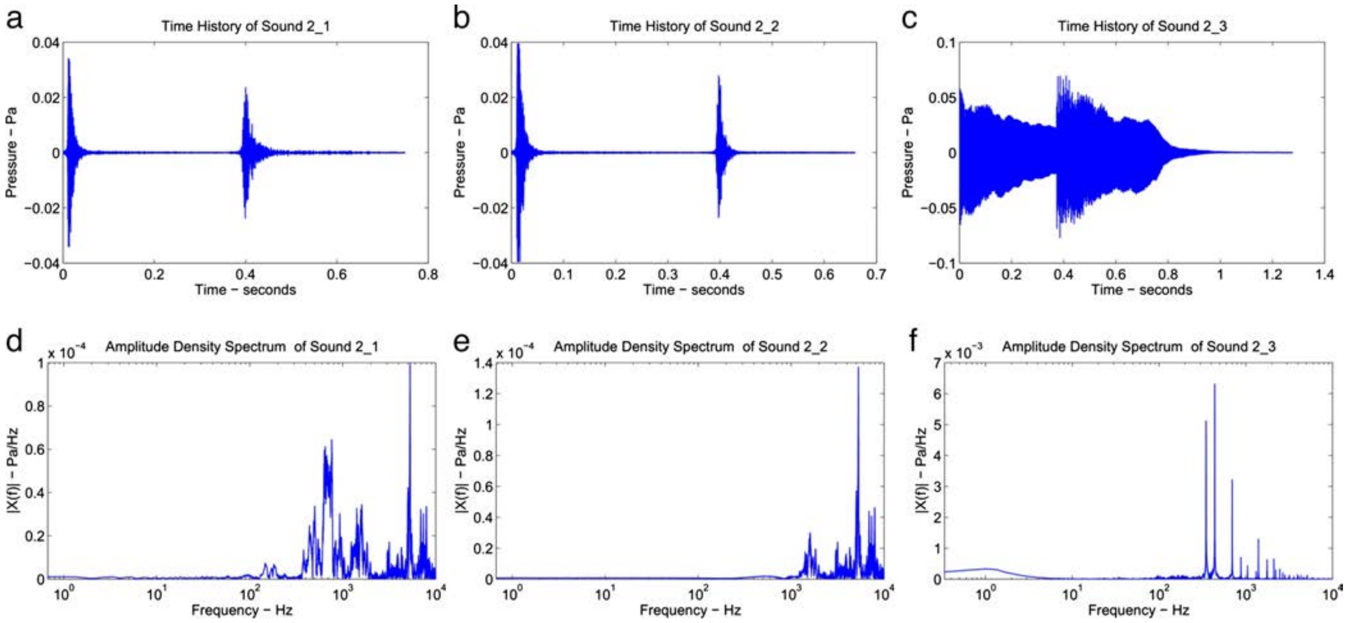

Fig. A2—The time histories and amplitude density spectrum of the sounds in Context 2 (turn signal). Subplots (a)–(c) are time histories of the "traditional," "between" and "very different" sounds, respectively. Subplots (d)–(f) are the associated amplitude density spectra.

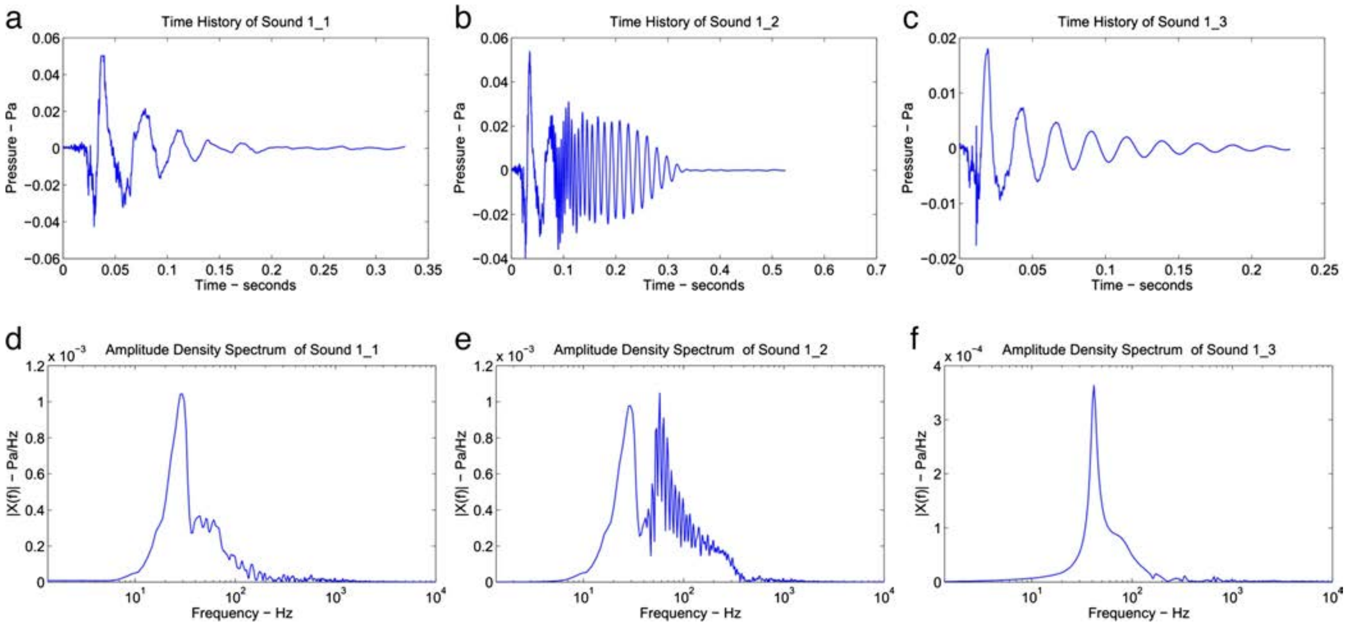

Fig. A3—The time histories and amplitude density spectrum of the sounds in Context 3 (put phone down). Subplots (a)–(c) are time histories of the "traditional," "between" and "very different" sounds, respectively. Subplots (d)–(f) are the associated amplitude density spectra.

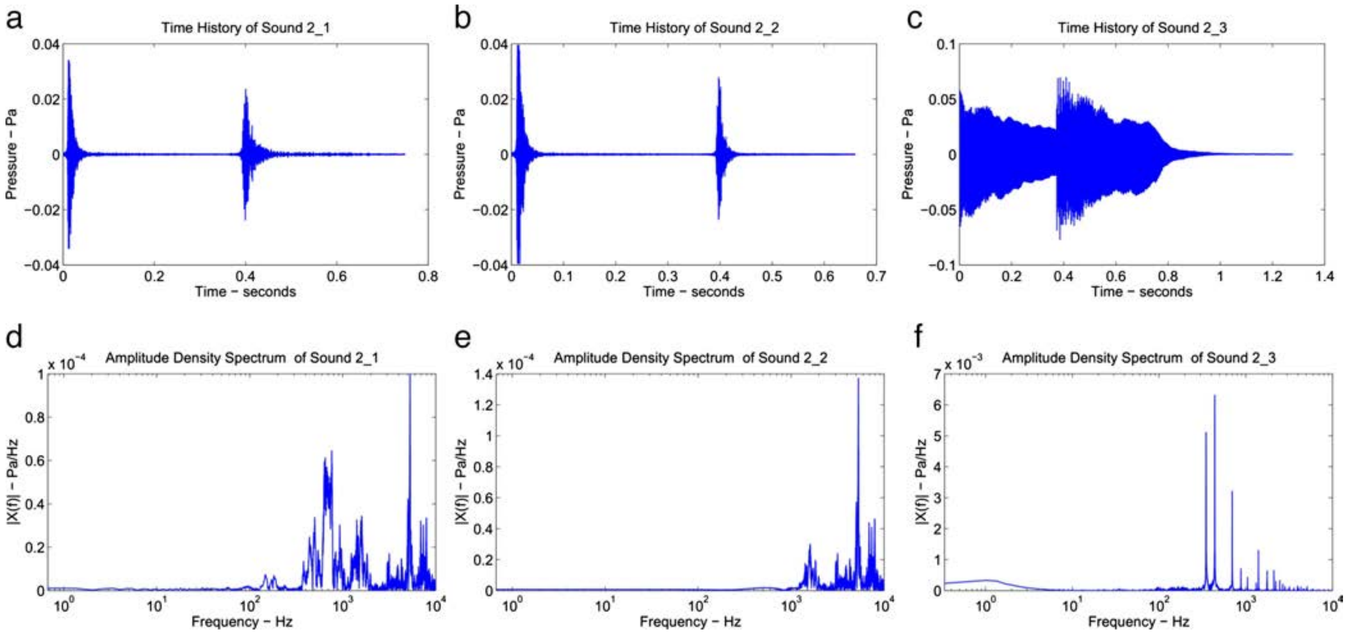

Fig. A4—The time histories and amplitude density spectrum of the sounds in Context 4 (camera click). Subplots (a)–(c) are time histories of the "traditional," "between" and "very different" sounds, respectively. Subplots (d)–(f) are the associated amplitude density spectra.

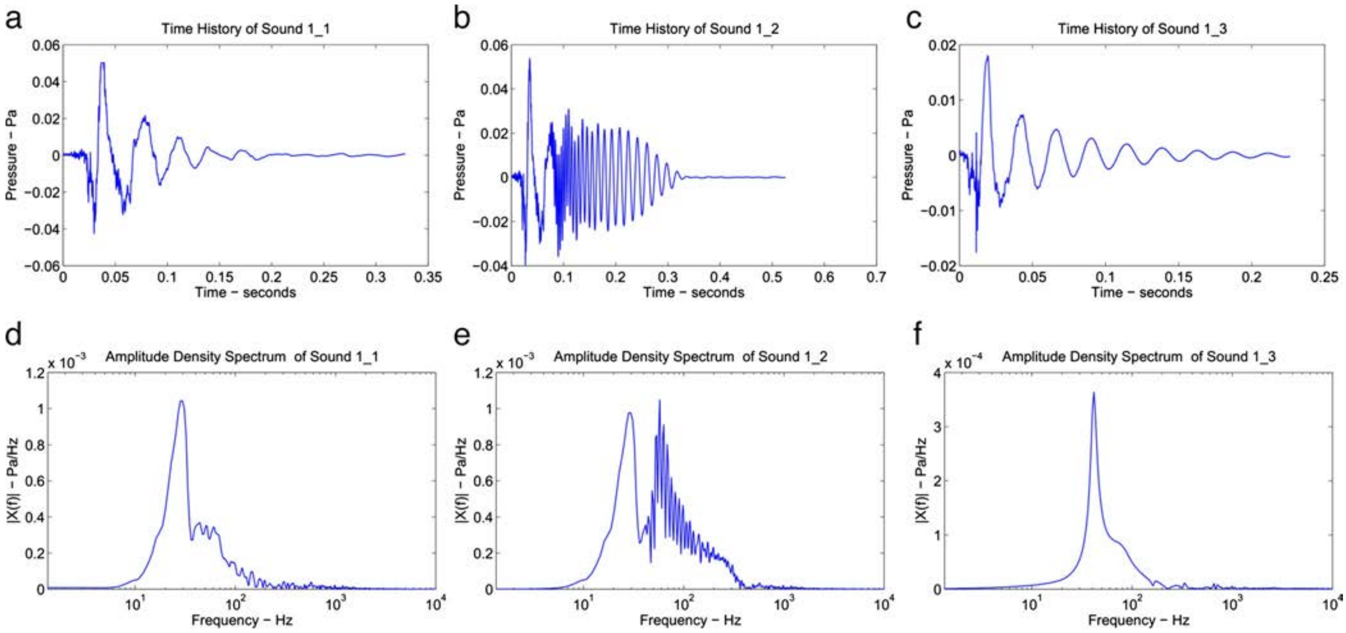

Fig. A5—The time histories and amplitude density spectrum of the sounds in Context 5 (car horn). Subplots (a)–(c) are time histories of the "traditional," "between" and "very different" sounds, respectively. Subplots (d)–(f) are the associated amplitude density spectra.

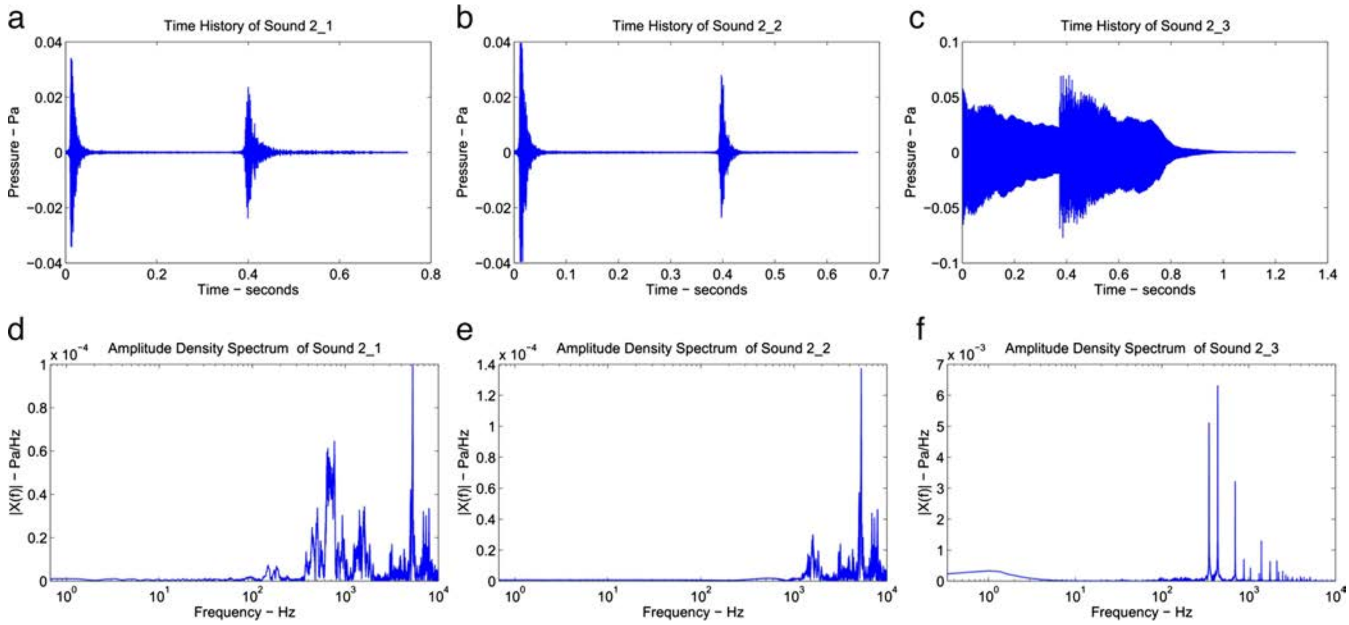

Fig. A6—The time histories and amplitude density spectrum of the sounds in Context 6 (windshield wipers). Subplots (a)–(c) are time histories of the "traditional," "between" and "very different" sounds, respectively. Subplots (d)–(f) are the associated amplitude density spectra.

estimated by using the discrete Fourier transform of the sampled time history scaled by the sampling interval ($|X(f)| \approx \Delta \cdot X_k$). The discrete Fourier transform was calculated by using the fft function in MATLAB.

8 REFERENCES

1. R. Van Egmond, "The Experience of Product Sounds", Chap. 3 in Product Experience, 69–89, Elsevier, San Diego, California, (2008).
2. T.G. Filippou, H. Fastl, S. Kuwano, S. Namba, S. Nakamura and H. Uchida, "Door sound and image of cars", Fortschritte Der Akustik (Progress of Acoustics), DAGA '03, Aachen, Germany, 306–307, (2003).
3. E. Parizet, E. Guyader and V. Nosulenko, "Analysis of car door closing sound quality", *Applied Acoustics*, **69**(1), 12–22, (2008).
4. J. Hyeonho, W. Seong, H. Lee, S. Kim, D. Park and Y. June, "A study on sound quality evaluation index of car door latch and improving sound quality by modifying door latch assembly design", *InterNoise14*, (2014).
5. D. Pick, D. Wang, R. Proctor and D. Patel, "Dead pedal and the perception of performance of a continuously variable transmission", SAE Technical Paper, 2005-01-1596, (2005).
6. J.H. Fleming, "Millennials Are Starting to Spend More", *Gallup Business Journal*, http://www.gallup. com/businessjournal/191837/millennials-starting-spend.aspx, (2016).
7. E.M. Nelson, "Millennials want to party with your brand but on their own terms", *DigitalNext*, http://adage.com/article/digitalnext/millennials-party-brand-terms/236444/, (2012).
8. A. Turley, "Millennial's, brands and social media", Bachelors Final Year Project Report, Dublin Business School, (2014).
9. L. Minnaar, "A manager's perspective on Generation Y in the hospitality industry in Finland", Bachelor's thesis for degree in Hotel, Restaurant and Tourism Management, Haaga-Helia University of Applied Sciences, (2014).
10. D. Oblinger, "Boomers, Gen-Xers, and Millennials: understanding the 'New Students'", *Educause Review*, **38**(4), 36–45, (2003).
11. R. Sweeney, "Millennial behaviors and demographics", Newark: New Jersey Institute of Technology, USA, https://certi.mst.edu/media/administrative/certi/documents/Article-Millennial-Behaviors.pdf, (2006).
12. T.H. Reisenwitz, and R. Iyer, "Differences in Generation X and Generation Y: implications for the organization and marketers", *The Marketing Management Journal*, **19**(2), 91–103, (2009).
13. D. Vastfjall, P. Larsson and M. Kleiner, "Cross-modal interaction in sound quality evaluation: some experiments using the virtual aircraft", *InterNoise02*, (2002).
14. S. Viollon, C. Lavandier, and C. Drake, "Influence of visual setting on sound ratings in an urban environment", *Applied Acoustics*, **63**(5), 493–511, (2002).
15. D. Menzel, N. Haufe and H. Fastl, "Colour-influences on loudness judgements", *20th International Congress on Acoustics ICA 2010*, (2010).
16. J. Yoshida, F. Volk, H. Fastl and G. Rigoll, "Influences of vehicle exterior images on sound quality ratings: German vs. Japanese drivers", *Acoustical Science and Technology*, **35**(6), 327–329, (2014).
17. P. Susini, O. Houix, N. Misdariis, B. Smith, and S. Langlois, "Instruction's effect on semantic scale ratings of interior car sounds", *Applied Acoustics*, **70**(3), 389–403, (2009).
18. E. Dimdam, "25 Most popular songs for the millennial generation", List25.com, Sports & Entertainment, http://list25.com/25-most-popular-songs-for-the-millennial-generation/5/, (2013).
19. R. Pratt, "Top 100 favorite films: according to a movie-watching millennial", AXS Entertainment/Arts & Entertainment/Movies, http://www.examiner.com/article/top-100-favorite-films-according-to-a-movie-watching-millennial, (2011).
20. A. Agresti, *Categorical Data Analysis*, Wiley Series in Probability and Statistics, John Wiley and Sons, Hoboken, NJ, USA, (2002).
21. D.R. Hunter, "MM algorithms for generalized Bradley-Terry models", *Annals of Statistics*, **31**(1), 384–406, (2004)
22. E. Zwicker and H. Fastl, *Psychoacoustics: Facts and Models*, 2nd Edition, Springer Science & Business Media, (2013).