Behav Res (2015) 47:1076–1084
DOI 10.3758/s13428-014-0525-4

Development of an auditory emotion recognition function using psychoacoustic parameters based on the International Affective Digitized Sounds

Youngimm Choi · Sungjun Lee · SungSoo Jung · In-Mook Choi · Yon-Kyu Park · Chobok Kim

Published online: 16 October 2014

Abstract The purpose of this study was to develop an auditory emotion recognition function that could determine the emotion caused by sounds coming from the environment in our daily life. For this purpose, sound stimuli from the International Affective Digitized Sounds (IADS-2), a standardized database of sounds intended to evoke emotion, were selected, and four psychoacoustic parameters (i.e., loudness, sharpness, roughness, and fluctuation strength) were extracted from the sounds. Also, by using an emotion adjective scale, 140 college students were tested to measure three basic emotions (happiness, sadness, and negativity). From this discriminant analysis to predict basic emotions from the psychoacoustic parameters of sound, a discriminant function with overall discriminant accuracy of 88.9 % was produced from training data. In order to validate the discriminant function, the same four psychoacoustic parameters were extracted from 46 sound stimuli collected from another database and substituted into the discriminant function. The results showed that an overall discriminant accuracy of 63.04 % was confirmed. Our findings provide the possibility that daily-life sounds, beyond voice and music, can be used in a human–machine interface.

Keywords Psychoacoustic parameters · Emotion recognition · Auditory emotion recognition · IADS-2 · Emotional adjectives

Y. Choi · S. Lee · S. Jung · I.-M. Choi · Y.-K. Park
Korea Research Institute of Standards and Science, Daejeon, South Korea

Y. Choi
Department of Psychology, Chungnam National University, Daejeon, South Korea

C. Kim (✉)
Department of Psychology, Kyungpook National University, 80 Daehak-ro, Buk-gu, Daegu 702-701, South Korea
e-mail: ckim@knu.ac.kr

The interaction between humans and machines to date has been machine-centric, whereby humans manipulate machines to operate them when needed. However, in the present day of human–machine interaction in ubiquitous environments, a more human-centric interaction is called for. In order to realize such human-centric interactivity, it is important for machines to actively and sensitively respond to our emotions and needs, and the core technologies essentially needed for this are emotion recognition technologies.

For example, intelligent cars that can recognize emotion could not only detect the driver's sentiment but also the driver's level of fatigue, sleepiness, and health status, thus assisting the driver to operate the vehicle safely (Katsis, Katertsidis, Ganiatsas, & Fotiadis, 2008). Specifically, they tried to detect a driver's emotional status including stress level, disappointment, and euphoria, using biological signals such as facial electromyograms, electrocardiogram, respiration, and electrodermal activity. Customer service call centers equipped with speech recognition technology can assign rhythms that suit each emotion and can even generate appropriate sound signals suitable for the users' emotions (Navas, Hernaez, & Iker, 2006). As such, emotion detection technologies will heighten the quality of human–machine interaction, and users will be able to interface with their computers in a more enjoyable and useful manner.

To date, studies on emotion recognition have been based on facial expressions (Zeng, Pantic, Roisman, & Huang, 2009), or focused on physiological indicators (Calvo & D'Mello, 2010) or voice recognition (Steidl, Levit, Batliner, Nöth, & Niemann, 2005). For example, Lee and colleagues (Lee, Narayanan, & Pieraccini, 2001) compared several classification rates by measuring speech of man versus women or changing feature extraction methods. Dellaert, Polzin, and Waibel (1996) also applied various feature extraction methods and analysis algorithms to speech stimuli. Among these, the field of research in emotion recognition through sound is

classified as psychoacoustics. In this field, people's sensitivities toward sound have typically been evaluated using psychoacoustic parameters including loudness, sharpness, roughness, fluctuation strength, and tonality (Zwicker, Flottorp, & Stevens, 1957). Specifically, recent studies that emphasized emotional responses to nonverbal voice or music have developed auditory emotional stimuli without verbal information. For example, Belin et al. (2008) validated 90 nonverbal emotional voice stimuli, The Montreal Affective Voices (MAV), and Paquette, Peretz, and Belin (2013) validated 80 short musical stimuli with happy, sad, fear, and neutral emotions. It would be also beneficial to use these stimuli in emotion recognition studies as well as emotion evaluation.

One of the reasons that psychoacoustics can be established as a distinct field of emotion recognition is that sound plays an important role in emotion recognition. For example, people can detect others' emotions from the tone of voice only even when they cannot hear the content of the message (Adolphs, Damasio, & Tranel, 2002), and can accurately interpret emotions such as anger, sadness, happiness, fear, and love through the voice even when listening to speech in foreign languages (Juslin & Laukka, 2003). Furthermore, many prior studies have substantiated that humans tend to adopt a multimodality percept using input from two or more sensory systems, rather than using input from a single system. For example, the accuracy of emotion recognition increases when using multimodality input with simultaneous visual and auditory stimuli, as compared to unimodality input, in which only facial expressions or sound stimuli are used separately (e.g., Castellano, Kessous, & Caridakis, 2008; Scherer & Ellgring, 2007).

Although a variety of sounds that evoke emotions do exist in daily life, psychoacoustic research on recognition of emotional response to sounds other than the human voice is very rare, and even those works deal with limited sources such as laughter or sneezes (Matos, Birring, Pavord, & Evans, 2006), or noises generated by heating and cooling devices or mobile phones (Kweon & Choi, 2009).

However, it is necessary to differentiate nonverbal sounds from human verbal sounds when examining emotional responses to those sounds. Specifically, it is well-known that the human voice is processed differently from other sounds in the brain. For example, Charest et al. (2009) found that voice and nonvoice sounds were rapidly discriminated during early perception processing in the brain. Also, Belin, Fillion-Bilodeau, and Gosselin (2008) suggested that speech with words or sentences yields in complex characteristics due to an interaction between emotional processing and verbal processing. Moreover, Paquette, Peretz, and Belin (2013) suggested that nonverbal voice is a relatively natural and universal means of human communication and less affected by cultures, resulting in stronger emotional responses.

Thus, it is important to examine emotional responses to any sounds from the environment. According to Gerhard (2003), the four categories of sounds include noise, natural sounds, artificial sounds, and speech. On the basis of this categorization, we included various types of sounds from the environment but excluded speech with words or sentences and music with melody or words. Accordingly, this study focused on the emotional responses to sounds from the environment, such as natural sounds, artificial sounds, nonverbal human sounds, and instrumental sounds.

Also, most emotion recognition studies have commonly applied a mechanical learning research methodology. To elaborate, the studies were conducted by extracting specific parameters that became apparent when an emotion was evoked (e.g., facial, voice or physiological indicators), training the data to predict an emotion by applying mathematical algorithms, and then validating those algorithm functions using new data. However, prior research has either remained at the level of simple sensitivity evaluations toward sounds or has employed data on other modalities other than the characteristics of the sounds themselves (e.g., Pal, Iyer, & Yantorno, 2006). Thus, it is thus difficult to classify those efforts as research on recognition of emotion caused purely by sound.

Upon this background, in the present study we sought to research the recognition of emotion using daily-life sounds, excluding the human voice and music. For this purpose, psychoacoustic parameters (Zwicker et al., 1957) were first extracted from various daily-life sounds. On the basis of those extracted parameters, and through an application of linear discriminant analysis (LDA) algorithms on the parameters, the authors aimed to deduce emotion recognition functions that would predict emotional responses to daily-life sounds (Fig. 1).

Method

Participants

A total of 140 students (70 females, 70 males), enrolled in Chungnam National University in Daejeon, South Korea, participated in the experiment. The average age was 23.56 (*SD* =2.08) and 21.34 (*SD* =1.99), respectively for males and females.

Step 1: Training data

Data sampling A total of 167 sound stimuli from IADS-2 (International Affective Digitized Sounds) created by Bradley and Lang (2007) were used to induce emotion associated with various sounds. IADS-2 is a standardized set of sound stimuli intended to induce emotion, and was developed for international usage in all areas of basic and applied psychology for

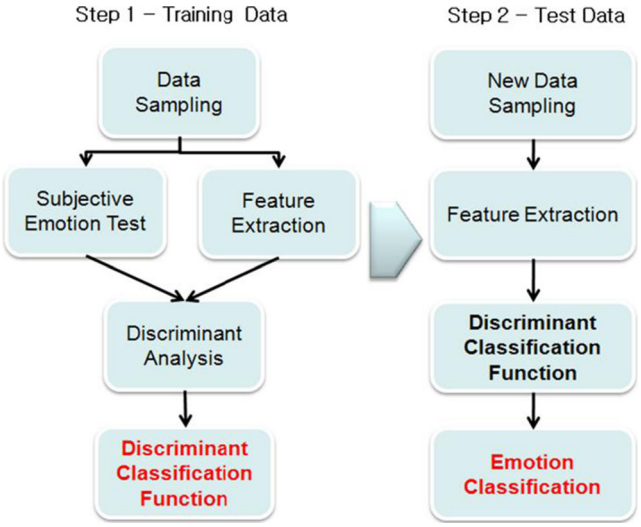

Fig. 1 A flowchart of emotion recognition model development.

better experimental control and comparison of emotion-related studies. The sound stimuli included in IADS-2 are distributed across two emotional dimensions (i.e., pleasure and arousal), and are controlled to be presented for 6 s.

Asking a single participant to evaluate his or her emotions for all 167 sounds could induce fatigue effects, and therefore only 60 sound samples among the total set of sound stimuli were introduced to each participant. In order to avoid a bias toward a specific emotion in a given set, the sound stimuli were selected using the following steps: as the IADS-2 sounds are distributed across two dimensions of pleasure and arousal, 15 sounds were randomly selected from each quadrant of the scatterplot (i.e., pleasure–high, arousal–high; pleasure–high, arousal–low; pleasure–low, arousal–low; and pleasure–low, arousal–high), resulting in three sets of stimuli, each consisting of 60 samples. Since there were a total of 167 stimuli, some of these were used twice and submitted with different sets of sounds.

Extraction of emotional adjectives from sounds Before conducting the experiment, open questionnaires were circulated to another 30 undergraduate students in order to collect frequently used emotional adjectives. They were asked to write adjectives to describe feelings of happiness, sadness, anger, fear, or disgust, and the 15 most frequent adjectives were selected from their responses. The participants, who were participated in the experiment, listened to the sounds presented by the experimenter and rated their emotional response levels on a Likert scale from 1 (*strongly disagree*) to 9 (*strongly agree*) against each of the 15 adjectives (see Table 1).

Design and procedure Participants were randomly assigned to groups of seven, resulting in a total of 20 groups participating in the experiment. Each group consisted of three or four females in order to provide control for the gender effect.

Table 1 Results of a factor analysis for emotional adjectives

	Factor		
Adjective	Negative	Happy	Sad
disgustful	**0.875**	–0.061	0.115
dislike	**0.871**	0.035	0.06
hatable	**0.869**	0.057	0.139
furious	**0.827**	–0.096	0.055
angry	**0.788**	–0.002	–0.097
enraged	**0.748**	0.072	–0.134
afraid	**0.697**	0.032	–0.195
fearful	**0.655**	–0.154	–0.102
scared	**0.62**	–0.142	–0.144
happy	0.027	**0.958**	0
pleasant	0.009	**0.949**	0.023
joyous	–0.082	**0.896**	–0.022
sorrowful	0.057	–0.004	**–0.819**
tearful	0.031	0.04	**–0.814**
sad	–0.035	–0.059	**–0.797**
Initial eigenvalues	7.586	1.995	1.817
% of variance (total 75.981 %)	50.573	13.297	12.112

The strongest factor for each adjective is in boldface.

For each group, an experimenter presented one of the three stimulus sets and each stimulus from the selected set in a random order. For generating the random orders of the stimulus sets and stimuli within a set, a full list of the sets and stimuli for each group was generated using the rand function implemented in Excel before conducting the experiment. The participant wrote down the number of the sound, listened carefully to the sound that was presented for 6 s until the end, and evaluated the emotion for each sound using the emotional adjective criteria based on a 9-point Likert scale. The stimuli were presented through the speakers (BR 1800, Britz, USA). In order to ensure that the relative magnitudes (in decibels) were identical at all of the participants' positions, the decibels of a sample sound (#102 in IADS-2, "cat") were measured by a decibel meter at all positions and were all the same, at 60 dB.

Step 2: Test data

Data sampling and procedure In order to validate the functions from the training session, a total of new 62 sound stimuli were obtained from the website www.findsounds.com. To label the newly collected sounds with emotions, 30 undergraduate students, who had not participated in the previous session, were asked to categorize these 62 sounds as one of the three emotions of happiness, sadness, and negative emotions (the degree of agreement among the evaluators was at Kappa = .74). Then the sounds for which

emotion classifications were mismatched between evaluators were excluded, resulting in 46 new sound stimuli (11 for happiness, seven for sadness, and 28 for negative emotions).

Results

Step 1: Training data

Subjective emotion test As a result of the factor analysis, in which the factors were extracted using the principal axis factoring method and then rotated using the direct oblimin method, three factors were extracted (Table 1). The first factor included "furious," "fearsome," "enraged," "scary," "disgusting," "chilling," "angering," and "revolting" and was named *negative emotions*. The second factor included "happy," "pleasing," and "joyous" and was named *happiness*, and the third factor that included "sad," "tearful," and "sorrowful" was named *sadness*.

Accordingly, each of the 167 sound stimuli was labeled as one of the three emotions of happiness, sadness, and negative on the basis of the factor in which they scored the highest. However, a sound was labeled as corresponding to a single emotion only when the average rating of that sound for that emotion was higher by ±1 standard deviation than the ratings for the other emotions. In other words, even if a sound scored highest in happiness, for example, if the score was within ±1 standard deviation from the scores for sadness or negative emotions, that sound was deemed as inducing duplicate emotions and was not labeled as happiness, and was thus discarded. Ultimately, 35 sounds that scored high for happiness, 14 sounds that scored high for sadness, and 33 sounds that scored high for negative emotions were used in the analysis, bringing the total number to 82.

Feature extraction From the 82 sounds selected, four factors—loudness, sharpness, roughness, and fluctuation strength—were extracted. Because these four physical characteristics showed a negative bias, they were subsequently log-transformed to create a normalized distribution.

Development of auditory emotion recognition function The discriminant analysis, which classified three emotions on the basis of four physical characteristics as the predictive variables, was repeatedly executed to train the data. Sounds were injected, the sounds that were not discriminated as one of the three emotions were removed, and different sounds were injected again in the repetitive execution of the discriminant analysis. Eventually, discriminant functions were generated from a total of 27 sounds for happiness, sadness, and negative emotions, with each of the categories having nine sounds, and the results are shown in Table 2.

The scatterplot of sounds according to their discriminant scores, derived from the canonical discriminant functions, is presented in Fig. 2. As a result of the discriminant analysis, the discriminant classifications of the functions classifying sounds corresponding to happiness, sadness, and negative emotions can be described as expressed in Points 1 to 3 below:

(1) $\text{Happy} = -442.450 + 705.127 \times L - 225.673 \times S + 87.168 \times R + 0.901 \times FS$,
(2) $\text{Sad} = -482.841 + 739.830 \times L - 255.673 \times S + 82.101 \times R + 2.987 \times FS$,
(3) $\text{Negative} = -555.843 + 794.830 \times L - 267.016 \times S + 95.006 \times R + 0.792 \times FS$,

where L =loudness, S =sharpness, R =roughness, and FS =fluctuation strength.

Applying the discriminant classifying function to the sounds used in the testing, seven of the nine happiness emotion sounds (77.8 %), eight of the nine sadness emotion sounds (88.9 %), and all nine of the negative emotions sounds (100 %) were accurately classified, resulting in 88.9 % classification accuracy overall (Table 3).

Since the coefficients for roughness and fluctuation strength in the functions were relatively lower than those for loudness and sharpness, it would be interesting to see how much each parameter or some of them contribute to the classification. In order to examine this, (1) each of the parameters was submitted to the analysis separately and (2) all of the parameters were submitted to the analysis in a stepwise manner in the order of loudness, sharpness, roughness, and fluctuation strength. The classification results showed that each of the parameters was lower than when all four parameters were submitted: loudness (63.0 %), sharpness (48.1 %), roughness (37.0 %), and fluctuation strength (40.7 %). Also, the stepwise analysis showed 63.0 % (loudness only), 70.4 % (loudness and sharpness; 7.3 % increase), 84.3 % (loudness, sharpness, and roughness; 13.9 % increase), and 88.9 % (loudness, sharpness, roughness, and fluctuation strength; 4.6 % increase).

Step 2: Test data

For this step, we executed discriminant analyses numerous times for a large number of stimuli included in IADS-2, and used only those sounds showing the highest discrimination accuracy to produce the function. Thus, the function would be susceptible to a data-dependent tendency. In order to exclude this possibility and to secure validity for the emotion recognition function produced above, the physical characteristics extracted from a new set of test data were substituted into the function generated using the training data.

Table 2 Canonical discriminant function coefficients and classification function coefficients

	Canonical Discriminant Function Coefficients		Classification Function Coefficients		
	Function 1	Function 2	Happy	Sad	Negative
Loudness	23.288	3.983	705.127	739.83	794.041
Sharpness	–10.647	6.026	–225.795	–255.673	–267.016
Roughness	2.141	4.991	87.168	82.101	95.006
Fluctuation Strength	–0.473	–1.644	0.901	2.987	–0.792
Constant	–29.698	–6.533	–442.45	–482.841	–555.843

Feature extraction For the 46 sounds, the four psychoacoustic characteristics for each were processed in order to produce psychoacoustic variables via the same method used for the stimuli from IADS-2, and the results are shown in Table 4.

Validation of auditory-emotion function For emotion recognition of the new sounds, the values of physical characteristics extracted from the new data were substituted into the emotion recognition functions of happiness, sadness, and negative emotions produced from the training data. The three emotion recognition function values were then computed, and the one with the highest score was discriminated as the emotion for that sound. For example, the values of the physical characteristics for the new-data Sound 1 shown in Table 4 are, respectively loudness =1.36, sharpness =0.49, roughness = –0.48, and fluctuation strength =0.32. Substituting these values into the functions for (1) happiness, (2) sadness, and (3) negative emotions, the function values are computed as, respectively, (1) happiness =364.331, (2) sadness =359.595, and (3) negative emotions =347.359:

(1) Happy=–442.450+705.127×1.36–225.673×0.49+87.168×(–0.48)+0.901×0.32=364.331,
(2) Sad=–482.841+739.830×1.36–255.67×0.49+82.101×(–0.48)+2.987×0.32=359.595,
(3) Negative=–555.843+794.830×1.36–267.016×0.49+95.006×(–0.48)+0.792×0.32=347.359.

Here, Sound 1 scored highest in the happiness emotion recognition function, and hence the emotion of Sound 1 was discriminated as "happiness." Sound 1 had been evaluated as a sound corresponding to "happiness" before the discrimination, and thus in this case the actual emotion and the emotion predicted through the discriminant classification function were consistent with each other. This means that the emotion recognition function produced from the training data

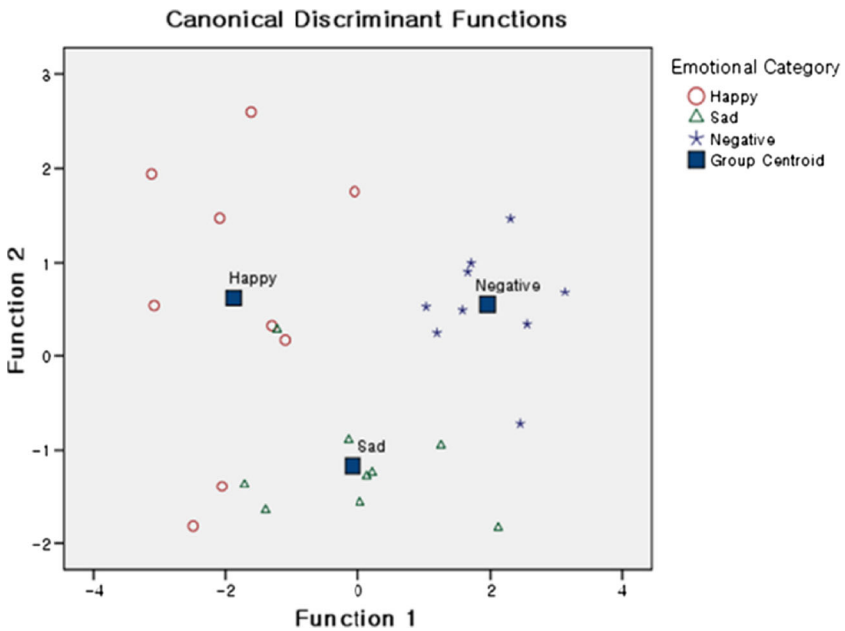

Fig. 2 A scatterplot for canonical discriminant functions from our training data.

Table 3 Classification results for our training data

	Emotion	Predicted Group Membership			Total
		Happy	Sad	Negative	
Count	Happy	7	2	0	9
	Sad	1	8	0	9
	Negative	0	0	9	9
%	Happy	77.8	22.2	0	100
	Sad	11.1	88.9	0	100
	Negative	0	0	100	100

88.9 % of the original grouped cases were correctly classified.

accurately discriminated the emotion of the sound belonging to the new data.

The emotion function values were computed for each of the 46 sounds belonging to the new data in order to predict the emotion for the sounds, and the results are shown in Table 4. The consistency frequency and overall accuracy between the actual emotion and the predicted emotion are presented in Table 5. As is shown in that table, nine of the 11 sounds corresponding to happiness were accurately predicted as such, three of the seven for sadness, and 17 of the 28 for negative emotions, resulting in an overall classification accuracy of 63.04 %.

Discussion

In this study, we aimed to derive an emotion recognition function that would predict emotional responses to daily-life sounds using psychoacoustic parameters extracted from various sound stimuli. By carrying out a factor analysis on the adjective measurement tool used to gauge the emotions, three emotions—happiness, sadness, and negative emotions (anger, fear, and disgust)—were extracted. This was in alignment with prior studies in which the emotions of anger, fear, and disgust were not readily distinguishable, or in which it was difficult to classify the three emotions through variances in physiological responses (Ekman, Levenson, & Friesen, 1983; Lang, 1994). According to Jack, Garrod, and Schyns (2014), in addition, emotional responses to facial expressions appear to be composed of fewer elementary categories, such as approach and avoidance or happy, sad, fear/surprise, and disgust/anger. This suggests that some of the basic emotions distinguished by Ekman include more complex emotions, and that it would be difficult to classify these complex emotions.

Through an evaluation of sounds selected from the IADS-2 database consisting of standardized emotion-inducing sound stimuli using an emotion measurement tool, we derived an emotion recognition function that exhibited classification accuracy of 88.9 % for the three emotions of happiness, sadness, and negative emotions. If we compare this to the results of prior studies, Dellaert et al. (1996) achieved an overall accuracy of 82 % when classifying sounds into the four emotions of happiness, sadness, anger, and fear using parameters extracted from their sounds, and Lee et al. (2001) reported a 77.5 % accuracy in classifying sounds into two emotional categories—negative and nonnegative—using an LDA algorithm. Therefore, the accuracy of classification in this study is believed to be acceptably high in comparison to prior works.

One might argue that six basic emotions should be considered in this study, not five emotions or three emotional factors. However, there is ongoing debate about how many basic emotions exist among the complex emotions. Usually the six basic emotions described by Ekman (1973) are assumed, but recent studies have suggested that there are five basic emotions, excluding "surprise" (e.g., Power, 2006). Moreover, it is important to consider the number of parameters to be used in predicting emotions, since statistical power would be decreased if the number of parameters to predict emotions were greater than the number of emotions. Thus, we used three emotional factors to be predicted by four psychoacoustic parameters.

Also, it is important to note that we used psychoacoustic parameters that have been widely used in various types of sounds (Fastl, 2005). The reason of using these was that various sound stimuli used in this study included natural sounds, artificial sounds, nonverbal human sounds, and instrumental sounds without melody. In order words, it is not likely to be acceptable to apply other parameters used in previous studies for speech (e.g., El Ayadi, Kamel, & Karray, 2011) or music (e.g., Fu, Lu, Ting, & Zhang, 2011) to our sound stimuli.

One of the reasons for the scarcity of psychoacoustic research to date on the subject of daily-life sounds, relative to research on the human voice or music, could be that a standardized emotion-inducing sounds database did not exist previously. The IADS-2 stimuli used in this study span a tremendously wide range of daily-life sounds, including human sounds other than speech (e.g., laughter, crying, sneezes, burps), natural sounds (e.g., waves, rain), animal sounds (e.g., cattle, cat, dog), mechanical sounds (e.g., car engine starting, trains, explosions), and household sounds (e.g., water flowing through a drain, knocking, a phone ringing). As such, these stimuli could be adopted beneficially by any research on emotional response through sound (Bradley & Lang, 2007). In particular, for this study, four acoustic characteristics—loudness, sharpness, roughness, and fluctuation strength—were extracted in order to predict three emotions. Although the validity of the parameters used for emotional cognizance in response to music and voice has been secured through repeated past studies, the empirical evidence for parameters

Table 4 Values of psychoacoustical parameters and discriminant classification results for 46 sounds from new data

Sound's No.	Sound's Name	Psychoacoustic Parameter				Classification Function Scores			Actual Group	Predicted Group
		L	S	R	FS	1	2	3		
1	crowd laugh	1.36	0.49	–0.48	0.32	364.33	359.60	347.36	1	1
2	baby's laugh	1.44	0.52	0.29	0.59	481.33	475.14	475.81		1
3	child laugh	1.18	0.46	–0.25	0.61	264.49	253.85	234.06		1
4	crowd	1.56	0.43	–0.46	–0.22	520.16	522.93	524.52		3
5	applause	1.44	0.45	–0.4	–0.27	436.22	433.81	429.63		1
6	applause2	1.31	0.48	–0.06	0.06	367.71	358.87	350.44		1
7	robin	1.19	0.56	–0.68	0.37	211.27	199.66	174.64		1
8	brook	1.77	0.57	0.02	–0.2	678.49	681.97	699.47		3
9	clap	1.46	0.5	0.6	0.5	526.89	520.23	526.56		1
10	coin	0.91	0.73	–0.12	0.34	24.23	–5.07	–39.86		1
11	soda	0.91	0.59	–0.89	0.26	–11.35	–32.74	–75.57		1
12	puppy	1.33	0.26	–0.95	0.45	354.26	358.01	340.20	2	2
13	baby cry	1.49	0.6	–0.2	0.34	455.59	450.70	447.80		1
14	woman cry	1.28	0.33	–0.03	0.64	383.56	379.22	369.06		1
15	couple sob	1.5	0.31	–0.22	0.61	526.62	531.41	531.06		2
16	cry	1.01	0.2	–0.57	0.4	175.24	167.65	138.27		1
17	choir	1.65	0.43	–0.66	–0.38	566.04	572.62	577.11		3
18	guitar	1.6	0.43	–0.76	–0.21	522.22	527.92	527.77		2
19	female scream	1.66	0.62	–0.68	–0.08	528.72	530.69	532.17	3	3
20	scream	1.66	0.62	–0.68	–0.08	528.72	530.69	532.17		3
21	buzzer	1.56	0.44	0.06	0.26	563.66	564.50	570.87		3
22	dog growl	1.15	0.12	0.19	0.44	358.31	354.20	342.97		1
23	bomb	1.54	0.44	–0.24	–0.11	523.08	523.97	526.78		3
24	siren	1.29	0.35	–0.6	–0.14	335.71	332.38	318.12		1
25	black bear	1.31	0.2	–0.15	0.14	423.16	423.31	416.59		2
26	tires skid	1.51	0.48	–0.16	0.26	500.20	499.22	499.58		1
27	car skid	1.24	0.42	–0.21	–0.12	318.66	309.57	296.77		1
28	bee	1.57	0.37	–0.45	–0.22	541.63	546.49	549.43		3
29	bees	1.6	0.56	–0.47	–0.07	518.28	518.91	520.50		3
30	plane crack	1.55	0.34	–0.31	–0.23	546.50	550.83	554.87		3
31	sneeze	1.35	0.55	0.09	0.57	393.64	384.40	377.35		1
32	typewrite	1.61	0.61	0.15	0.37	568.48	565.75	573.64		3
33	jackhammer	1.81	0.74	–0.43	–0.17	629.11	631.24	643.06		3
34	shovel	1.32	0.6	0.05	0.71	357.84	346.56	336.27		1
35	helicopter	1.65	0.68	0.08	–0.11	574.34	570.26	580.44		3
36	crash	1.84	0.71	–0.16	0.1	680.81	684.08	700.33		3
37	bomb explosion	1.62	0.38	0.07	0.35	620.47	625.32	635.41		3
38	belch	1.58	0.41	0.17	0.14	594.02	595.64	605.31		3
39	alarm	1.46	0.67	0.29	0.56	461.54	451.49	451.66		1
40	brush teeth	1.53	0.71	0.17	0.31	491.18	482.45	485.36		1
41	glass break	1.5	0.71	0.49	0.76	498.32	487.88	491.59		1
42	crash	1.54	0.42	–0.14	0.09	536.49	537.89	541.46		3
43	buzzer	1.6	0.4	–0.5	0.02	551.87	557.63	560.30		3
44	nose blow	1.72	0.37	–0.39	0.14	652.96	663.47	673.95		3
45	car drive	1.7	0.58	–0.41	–0.27	589.32	592.11	600.42		3
46	static electronic	1.49	0.67	0.28	–0.32	481.03	470.24	475.23		1

1 =happy, 2 =sad, 3 =negative; L =loudness, S =sharpness, R =roughness, FS =fluctuation strength

Table 5 Classification validation results for the new-data sounds

	Emotion	Predicted Group Membership			Total
		Happy	Sad	Negative	
Count	Happy	9	0	2	11
	Sad	3	3	1	7
	Negative	10	1	17	28
%	Happy	81.8	0	18.2	100
	Sad	42.9	42.9	14.3	100
	Negative	35.7	3.6	60.7	100

63.04 % of the original grouped cases were correctly classified.

to be used for emotional response to other sounds has been extremely limited. Since this research used psychoacoustic parameters to predict people's emotional responses to various sounds, the results could be utilized as empirical evidence for supporting future research.

By applying the emotion recognition function derived from the training data to the test data, 63.04 % recognition accuracy was achieved. This is approximately 20 % lower than the accuracy achieved with the training data set. This decrease could be attributed to several reasons. First, the data used for the training were derived from the IADS-2 database, but the test data set were collected from the Internet in order to secure ecological validity. In other words, the stimuli included in the test data were, in contrast to the IADS-2 stimuli, uncontrolled in terms of recording type (stereo or mono), bit rate, frequency (in hertz), and so forth, and it is possible that this affected the discriminant accuracy.

Second, in this study the accuracy for recognizing sadness was found to be lower than that for the other emotions. This we attributed to the smaller number of stimuli expected to induce sadness, as compared to the stimuli for the other emotions (i.e., in the training data, 14 sounds were expected to induce sadness, as compared to 35 for happiness and 35 for negative emotions, and the test data contained seven sounds for sadness versus 11 for happiness and 28 for negative emotions). As such, the training process was insufficient to derive a function for recognizing sadness, and we assume that there were limitations to applying the function on the test data, as well.

Third, all emotional recognition studies, including the present study, face a common limitation in that they cannot consider the cognitive factors of emotional determination. In other words, according to Schachter and Singer's (1962) two-factor theory of emotion, although a person experiences physiological excitation when feeling emotion, since all emotions induce similar excitations, humans end up making a cognitive evaluation with regard to the contextual circumstances that evoked such physiological excitations (Kalat & Shiota, 2007). For example, for the sound of laughter, people differentiate the emotion caused by the laughter of joy versus the laughter of criticism or derision. However, because the physical parameters extracted from all instances of laughter are indistinguishable, an emotion recognition function using physical parameters will determine both types of laughter as inducing identical emotional responses. Therefore, emotion recognition using physical characteristics only, in the absence of human cognitive judgment, has limitations.

Despite these limits, our findings provide the possibility that daily-life sounds can be used in technology for emotional recognition. If technology for recognizing emotional response to sound using the present results were incorporated with existing emotional recognition research using facial expression, voice, or physiological indicators, we could look forward to human–machine interaction that was more similar to human emotional processing, as has been shown by the results of recent emotion recognition research based on multiple modalities.

Author note This research was supported by the Converging Research Center Program funded by the Ministry of Education, Science and Technology (No. 2012K001326) and the Next Generation Fire Protection & Safety Core Technology Development Program funded by the National Emergency Management Agency (No. NEMA-NG-2014-53).

References

Adolphs, R., Damasio, H., & Tranel, D. (2002). Neural systems for recognition of emotional prosody: A 3-D lesion study. *Emotion, 2,* 23–51. doi:10.1037/1528-3542.2.1.23

Belin, P., Fillion-Bilodeau, S., & Gosselin, F. (2008). The Montreal Affective Voices: A validated set of nonverbal affect bursts for research on auditory affective processing. *Behavior Research Methods, 40,* 531–539. doi:10.3758/BRM.40.2.531

Bradley, M. M., & Lang, P. J. (2007). *International Affective Digitized Sounds (2nd Edition; IADS-2): Affective ratings of sounds and instruction manual* (Technical Report No. B-3). Gainesville, FL: University of Florida, NIMH Center for the Study of Emotion and Attention.

Calvo, R. A., & D'Mello, S. (2010). Affect detection: An interdisciplinary review of models, methods, and their applications. *IEEE Transactions on Affective Computing, 1,* 18–37.

Castellano, G., Kessous, L., & Caridakis, G. (2008). Emotion recognition through multiple modalities: Face, body gesture, speech. In C. Peter & R. Beale (Eds.), *Affect and emotion in human–computer interaction* (Vol. 4868, pp. 92–103). Berlin, Germany: Springer.

Charest, I., Pernet, C. R., Rousselet, G. A., Quiñones, I., Latinus, M., Fillion-Bilodeau, S., … Belin, P. (2009). Electrophysiological evidence for an early processing of human voices. *BMC Neuroscience, 10*, 127–138. doi:10.1186/1471-2202-10-127

Dellaert, F., Polzin, T., & Waibel, A. (1996, October). *Recognizing emotion in speech.* Paper presented at the Fourth International Conference on Spoken Language (ICSLP 96), Philadelphia, PA.

Ekman, P. (1973). Cross-cultural studies of facial expression. In P. Ekman (Ed.), *Darwin and facial expression: A century of research in review* (pp. 169–222). New York, NY: Academic Press.

Ekman, P., Levenson, R. W., & Friesen, W. V. (1983). Autonomic nervous system activity distinguishes among emotions. *Science, 221,* 1208–1210.

El Ayadi, M., Kamel, M. S., & Karray, F. (2011). Survey on speech emotion recognition: Features, classification schemes, and databases. *Pattern Recognition, 44,* 572–587. doi:10.1016/j.patcog.2010.09.020

Fastl, H. (2005). Psycho-acoustics and sound quality. In J. Blauert (Ed.), *Communication acoustics* (pp. 139–162). Berlin, Germany: Springer.

Fu, Z., Lu, G., Ting, K. M., & Zhang, D. (2011). A survey of audio-based music classification and annotation. *IEEE Transactions on Multimedia, 13,* 303–319.

Gerhard, D. (2003). *Audio signal classification: History and current techniques* [Working paper]. Department of Computer Science, University of Regina, Regina, Saskatchewan, Canada.

Jack, R. E., Garrod, O. G., & Schyns, P. G. (2014). Dynamic facial expressions of emotion transmit an evolving hierarchy of signals over time. *Current Biology, 24,* 187–192. doi:10.1016/j.cub.2013.11.064

Juslin, P. N., & Laukka, P. (2003). Communication of emotions in vocal expression and music performance: Different channels, same code? *Psychological Bulletin, 129,* 770–814. doi:10.1037/0033-2909.129.5.770

Kalat, J. W., & Shiota, M. N. (2007). *Emotion*. Belmont, CA: Thomson Wadsworth.

Katsis, C. D., Katertsidis, N., Ganiatsas, G., & Fotiadis, D. I. (2008). Toward emotion recognition in car-racing drivers: A biosignal processing approach. *IEEE Transactions on Systems, Man, and Cybernetics Part A: Systems and Humans, 38,* 502–512.

Kweon, O., & Choi, J. (2009). Preliminary study on human sensibility evaluation of ringtone in mobile phone. *Koean Journal of the Science of Emotion and Sensibility, 12,* 403–410.

Lang, P. J. (1994). The varieties of emotional experience: A meditation on James–Lange theory. *Psychological Review, 101,* 211–221. doi:10.1037/0033-295X.101.2.211

Lee, C. M., Narayanan, S., & Pieraccini, R. (2001, December). *Recognition of negative emotions from the speech signal.* Paper presented at the IEEE Workshop on Automatic Speech Recognition and Understanding (ASRU '01), Madonna di Campiglio, Italy. doi:10.1109/ASRU.2001.1034632

Matos, S., Birring, S. S., Pavord, I. D., & Evans, H. (2006). Detection of cough signals in continuous audio recordings using hidden Markov models. *IEEE Transactions on Biomedical Engineering, 53,* 1078–1083.

Navas, E., Hernaez, I., & Iker, L. (2006). An objective and subjective study of the role of semantics and prosodic features in building corpora for emotional TTS. *IEEE Transactions on Audio, Speech and Language Processing, 14,* 1117–1127.

Pal, P., Iyer, A. N., & Yantorno, R. E. (2006, May). *Emotion detection from infant facial expressions and cries.* Paper presented at the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2006), Toulouse, France.

Paquette, S., Peretz, I., & Belin, P. (2013). The "Musical Emotional Bursts": A validated set of musical affect bursts to investigate auditory affective processing. *Frontiers in Psychology, 4,* 509. doi:10.3389/fpsyg.2013.00509

Power, M. J. (2006). The structure of emotion: An empirical comparison of six models. *Cognition and Emotion, 20*(5), 694–713.

Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological review, 69*(5), 379–399.

Scherer, K. R., & Ellgring, H. (2007). Multimodal expression of emotion: Affect programs or componential appraisal patterns? *Emotion, 7,* 158–171.

Steidl, S., Levit, M., Batliner, A., Nöth, E., & Niemann, H. (2005, March). *"Of all things the measure is man": Automatic classification of emotions and inter-labeler consistency.* Paper presented at ICASSP 2005, Philadelphia, PA.

Zeng, Z., Pantic, M., Roisman, G. I., & Huang, T. S. (2009). A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 31,* 39–58.

Zwicker, E., Flottorp, G., & Stevens, S. S. (1957). Critical band width in loudness summation. *Journal of the Acoustical Society of America, 29,* 548–557. doi:10.1121/1.1908963