PERCEPTION OF EMOTIONS IN KNOCKING SOUNDS: AN EVALUATION STUDY

Malcolm Houel, Abhilash Arun
Alfred Berg, Alessandro Iop
KTH Royal Institute of Technology, Sweden
{mhouel,aarun,alfrberg,aiop}@kth.se

Adrián Barahona-Ríos
Computer Science,
University of York, UK
ajbr501@york.ac.uk

Sandra Pauletto
KTH Royal Institute of
Technology, Sweden
pauletto@kth.se

ABSTRACT

Knocking sounds are highly meaningful everyday sounds. There exist many ways of knocking, expressing important information about the state of the person knocking and their relationship with the other side of the door. In media production, knocking sounds are important storytelling devices: they allow transitions to new scenes and create expectations in the audience. Despite this important role, knocking sounds have rarely been the focus of research. In this study, we create a data set of knocking actions performed with different emotional intentions. We then verify, through a listening test, whether these emotional intentions are perceived through listening to sound alone. Finally, we perform an acoustic analysis of the experimental data set to identify whether emotion-specific acoustic patterns emerge. The results show that emotional intentions are correctly perceived for some emotions. Additionally, the emerging emotion-specific acoustic patterns confirm, at least in part, findings from previous research in speech and music performance.

1. INTRODUCTION

Knocking sounds are extremely common in our everyday lives. They are also an important storytelling tool in film, games, theatre, VR, and other media. In this context they are often used as sound transitions to new scenes or to new twists and turns in the story. When we knock, we communicate with the people behind the door. We are often "asking" whether we can enter, if we can open the door, or if they can open the door. The way we knock says whether we feel entitled to enter, or if we are worried about entering. There are even knocking patterns, developed within cultures, which are commonly understood in those contexts. Knocking sounds therefore tell us about the person knocking on the door, about the relationship between the person knocking and the person behind the door, as well as create expectations in the listener. In everyday life, and in audiovisual media, there is a lot of contextual information that contributes to our ability to interpret the meaning and the emotional content of a knocking sound. This study aims to explore to what extent we can perceive an intended emotion from the knocking sound alone. Furthermore, following an acoustic analysis of the knocking sounds utilised in this study, we hypothesise which acoustic cues might contribute to the perception of emotions in these sounds.

Copyright: © 2020 Malcolm Houel, Abhilash Arun et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

2. BACKGROUND

2.1 Basic Emotions

The emotions studied in this research are part of a group called basic emotions (fear, anger, sadness, happiness, surprise and disgust). Basic emotions are considered to be fundamentally different from each other, and to be the root of more complex or compounded emotions [1]. Anger, for instance, could produce hot fury or cold annoyance. Additionally, it is thought that basic emotions have evolved from the need to deal with fundamental life tasks: "the primary function of emotion is to mobilise the organism to deal quickly with important interpersonal encounters, prepared to do so by what types of activity have been adaptive in the past" [2, p. 171]. Many studies that aim to unveil how we perceive emotions through sound (for example through speech or music) have focused, at least initially, on basic emotions as the clear differences in the nature of the emotions might produce quite marked responses that can form the basis for more nuanced studies.

2.2 Emotions and Sound

Research on speech and emotions, and music and emotions, has a relatively long history [3, 4]. This research has demonstrated that different acoustic cues correlate to the perception of different emotions both in speech and music. Additionally, researchers have found that there might be a close relationship between vocal expression of emotions and musical expression of emotions, which is consistent with an evolutionary perspective on vocal expression of emotions [5]. Research on the perception of emotions in everyday sounds is, on the other hand, very limited. Research has shown that humans can react emotionally to all sounds, even to meaningless sounds [6], that we react to acoustic stimuli emotionally similarly to how we react to images [7], and that we use emotions to categorise sounds [8]. Knowledge about emotion perception in everyday sounds can greatly contribute to the development of methods for automatic assessment of evoked emotions via, for example, transfer learning [9], something which is of great interest for sound design and au-

ditory displays. From our everyday experiences, it is clear that everyday sounds communicate emotions: we have all heard a timid knock, an angry door slam, fearful footsteps, happy fingers tapping on a table, etc. Despite the ubiquity of these sounds, little research has been carried out on how this information is encoded in the sound signal alone. Most studies on everyday sounds have focused on how we group sounds together or recognise the physical causes of the sounds [10–14]. Alarms and footsteps are among the most studied everyday sounds. Alarm sounds are often studied in relation to their effectiveness in communicating urgency in high pressure environments such as hospitals [15, 16], while footsteps are studied in relation to how well they portray the the characteristics of a person (gender, weight, gait, etc.) [17, 18] or how they contribute to presence [19], for example, in VR. Giordano *et al.* [20] have also suggested that both the expression and recognition of emotion in music might, at least in part, rely on knowledge about the sounds of expressive body movements, such as walking sounds. Research in this area [21] has shown that emotional transforms can be derived from the study of expressive movement, which can, for example, generate emotional animations of human figures from Neutral ones, additionally research has shown [22] that an observer can recognise positive and negative emotions in a knocking movement. In regard to emotions and knocking sounds without a visible accompanying gesture, one study exists [23] by colleagues at our department, which unfortunately is only partially published. The aim of the study was to investigate how temporal and dynamic cues in knocking sounds can communicate emotions, just like in expressive musical performances. Acoustic data as well as visual and gesture (motion capture) data were gathered. Results confirmed that same strategies are used in both expressive everyday body gestures and expressive music performance, and that listeners were, to a large extent, able to perceive emotions in the knocking sounds. We used some of the information available from this study as a starting point for the study presented here.

3. METHOD

The present study followed the following four steps:

- Recording knocking sequences with specific emotional intentions
- Designing a listening test to evaluate whether the knocking sequences successfully conveyed the intended emotion
- Analysing the acoustic features of the knocking sequences by emotion
- Assessing whether correlations exist between acoustic features and the emotional categorisation of the knocking sequences

Emotion	Scenario
Happiness	Telling a flatmate you won a prize
Sadness	Telling a friend someone passed away or they are moving to another country
Anger	Telling a flatmate for the fourth time to turn down the very loud music
Fear	Alerting a neighbour of a risk (e.g. fire)
Neutral	Parcel delivery

Table 1. Emotions scenarios for recording

3.1 Recording Knocking Sequences with Emotional Intentions

The recording session was carried out in a single day in a quiet room (Multistudio, Media Technology and Interaction Design Division, KTH) sized 27 m^2. The knocking sequences were performed on an MDF door measuring 203.7 cm $\times$ 72.6 cm. The recording was done using a Røde NTG2 super cardioid shotgun microphone with 76dB SPL signal/noise ratio along with a Zoom H4n Pro recorder (sample rate 48 kHz; bit depth 32 bit). Three men and three women (average age = 23.16 years; standard deviation = 1.72) from different countries (Germany, India, Italy, Serbia, USA) performed the knocking sequences. Each participant was asked to specify an area on the door where they felt comfortable knocking, and the microphone was then placed about 35 cm from the center of the specified area (see Figure 1). After a test recording, participants were asked to knock according to given scenarios, which aimed to evoke a specific emotion (see Table 1).

Every participant recorded a total of 20 knocking actions for each of the five scenarios for a total of 600 sequences. All instructions were given vocally. The participants received no payment for their contribution.

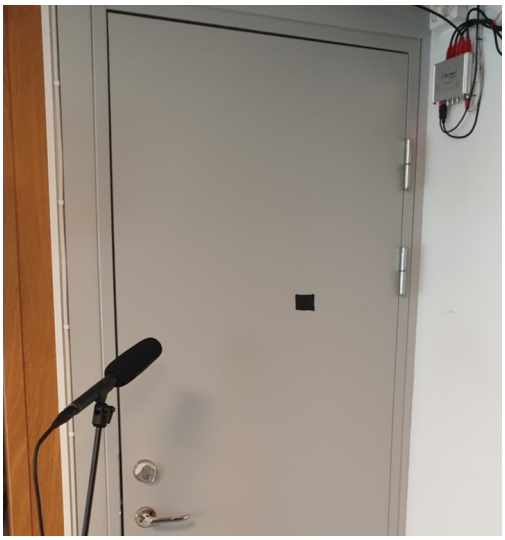

Figure 1. Recording set up

3.2 Listening Test

In order to keep the duration of the perception test within a reasonable length (10-15 minutes), five recordings were randomly selected for each emotion and each performer. This meant that, for each emotion, one of the participants (not always the same) was not represented. These selection criteria aimed to ensure that the diversity of performance was maintained in the reduced data set. In total, the test contained 25 sounds, which can be found here `https://tinyurl.com/y8qhp5b4`. We verified that this reduced data set was representative of the whole data set (see Figure 2).

The test was carried out online using a survey tool called SoGoSurvey [24]. We chose to carry out the test online rather than in a controlled environment in order to reach a larger number of participants, and also in order to use a more ecologically valid environment (house, office, etc.) where we would normally hear these kinds of sounds both in real life or from media. Participants were sent an email with a link to the survey which contained a short introduction and instructions on how to do the test. Participants were asked to report age and gender. They were also encouraged to use headphones and adjust the volume to a comfortable level. The 25 knocking sequences were presented in a different random order to each participant. After listening to each knocking action, participants were asked to select one of five emotions. The emotions were displayed as a horizontal line of radio buttons. The order of the radio buttons was randomised for every knocking action.

3.3 Acoustic Analysis

An acoustic analysis of the whole data set, as well as the 25 knocking actions present in the test, was carried out. We focused mostly on temporal characteristics (action duration, knocking rate, number of knocks) and relative characteristics (loudness slope and regularity of the knocks). Spectral characteristics were considered highly dependent on the door material and therefore were not examined in this study. The following parameters were considered:

- *Action duration*: Length of the knocking action. This is defined as the time passed from the first knock onset to the last knock decay.

- *Number of knocks per action*: The number of knocks was retrieved by counting the number of onsets detected in each audio file.

- *Knocking rate*: The knocking rate was retrieved by dividing the number of knocks in an action by the total time of the action. This feature is applied only to actions with two or more knocks.

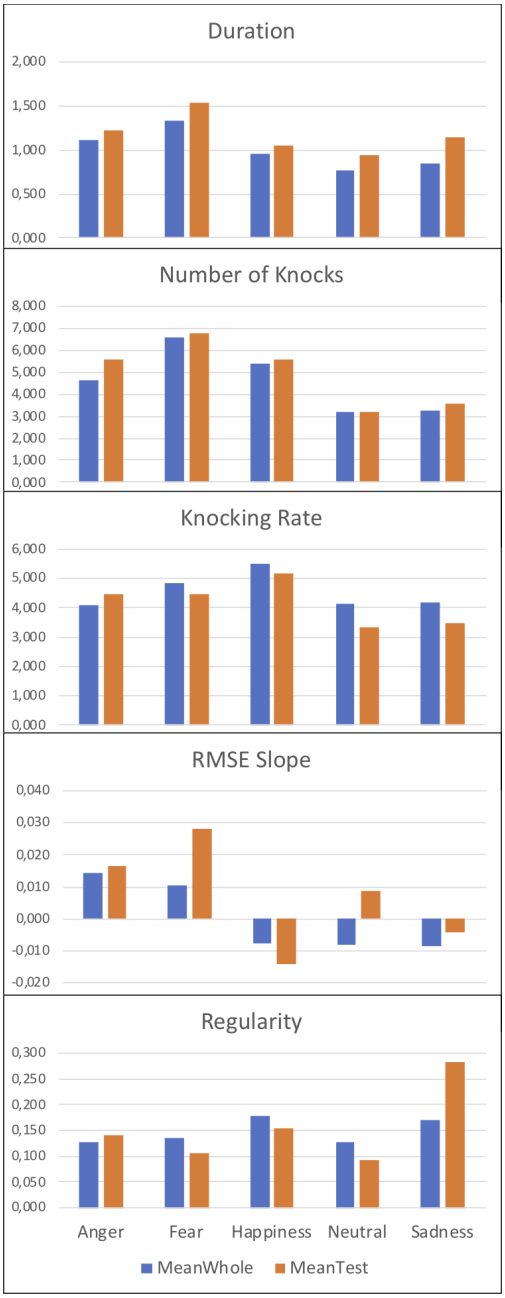

Figure 2. Comparison between whole dataset and the reduced dataset. The emotion labels' order (Anger, Fear, Happiness, Neutral, Sadness) remain constant for all acoustic characteristics.

- *Knocking regularity*: The knocking regularity measures how regular the knocks are in an action. Actions where the knocks are performed on a steady pace will show a higher regularity. To extract this feature, we calculated the inter-onset interval (IOI) of each action and computed the coefficient of variation (the standard deviation divided by the mean). Irregular actions will have higher coefficient of vari-

ation. This feature is applied only to actions with more than 2 knocks.

- *Root-mean-square energy (RMSE) slope*: This retrieves the energy pattern (crescendo or decrescendo) of an action. We calculated the root- mean-square energy of each individual knock and applied a linear regression to each action. The slope of the fitted line determines the energy crescendo (positive values) or decrescendo (negative values) of the action. An example of this feature applied to an individual action can be seen in Figure 3. This feature is applied only to actions with two or more knocks.

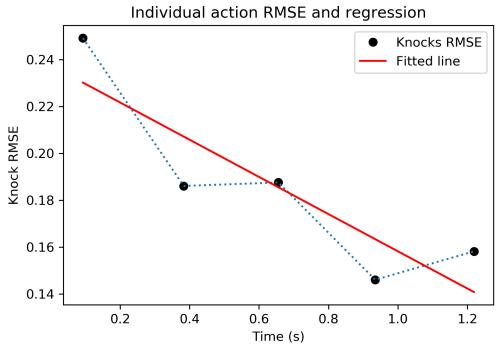

Figure 3. Example of the root-mean-square energy feature. The X axis represents knocking action duration in seconds. The Y axis represents the RMSE of the individual knocks. The individual knock positions in the action are represented by the black dots. The slope of the fitted line (in red) is negative, therefore the action has a decrescendo energy pattern.

These features were utilised by the authors in a parallel study on the synthesis of knocking actions [25] in order to allow future comparison of the results.

4. RESULTS

4.1 Listening Test

One hundred participants took part in the survey, however two participants did not complete the test and therefore their data was eliminated. The online survey was completed by 98 participants (59 male, 39 female, 0 other). The majority of the subjects had an age between 18-24 (number of subjects (age range): $1(<18); 65(18-24); 24(25-39); 6(40-60); 2(>60)$). The data was analysed with a Chi-square test for independence using SPSS. Overall, the relation between the knocking actions' emotional intention and the subjects' perceived emotion was significant, $\chi 2(96, N = 2450) = 2277.58, p < 0.001$. Comparisons between perceived emotions results were performed using a z-test and Bonferroni adjusted p-value.

Intended Emotion	Perceived Emotion/s (Percentage)
Anger	Anger (47.1%)
Fear	Anger (49%); Fear (23.3%)
Happiness	Happiness (35.1%)
Neutral	Neutral (49%); Sadness (22.9%)
Sadness	Sadness (40%)

Table 2. Statistically significant perceived emotion results. If two emotions are stated for one intended emotion, it means that there is no statistically significant difference between the two perceived emotions

Focusing on the specific knocking actions: two Anger, one Happiness, two Sadness, and two Neutral actions were perceived correctly more than 50% of the time. On the other hand, three Fear actions were perceived as Anger more than 50% of the time. One Sadness action was perceived as Neutral more than 50% of the time.

4.2 Acoustic Analysis

In this section we present the acoustic analysis of the listening test sounds. When appropriate, similarities or differences between the findings in this study and findings for cross-modal patterns of acoustic cues in vocal expression and music performance for discrete emotions by Juslin and Laukka summarised in Table 11, p. 802 [5] will be highlighted.

4.2.1 Duration of Knocking Action

- On average, Fear sounds are longer than knocking actions with other emotions (minimum difference 0.31 sec).
- On average, Happiness and Neutral actions are shorter.

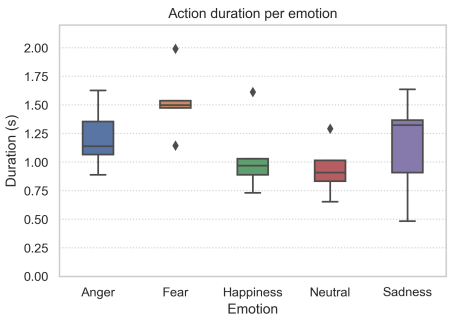

Figure 4. Duration of knocking action

4.2.2 Number of Knocks per Action

- On average, Fear presents the highest number of knocks per action (6.8).
- Anger and Happiness have the same average (5.6).
- Sadness and Neutral have similarly lower averages (3.6 and 3.2).

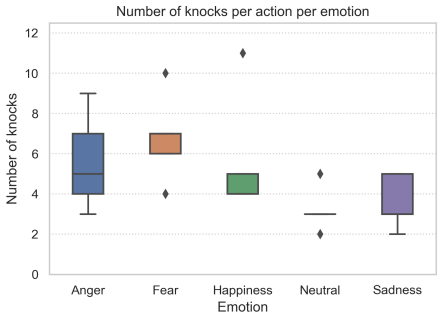

Figure 5. Number of knocks

4.2.3 *Knocking rate*

- Happiness knocking actions have the highest average knocking rate above 5 knocks per second (5.1 knocks per second).
- Anger and Fear present a similar average knocking rate above 4 knocks per sec.
- Sadness and Neutral present a lower average knocking rate (below 4 knocks per sec).
- The average knocking rate for Sadness has the highest variability (SD=1.34).

This result is consistent with results by [5] where fast rate/tempo were found to be acoustic cues for Happiness, Anger and Fear, and low rate/tempo for Sadness.

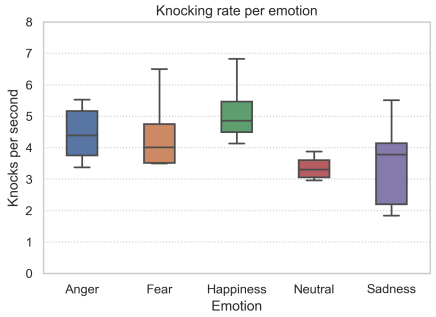

Figure 6. Knocking rate

4.2.4 *Knocking Regularity*

- On average, Sadness actions are the least regular of all knocking actions.
- Other actions are similarly regular.

The result for Happiness matches results in [5]. However, results for Anger, Fear and Sadness are not consistent with [5] as Juslin and Laukka find that Anger and Fear sounds are characterised by irregularity, while Sadness sounds are more regular.

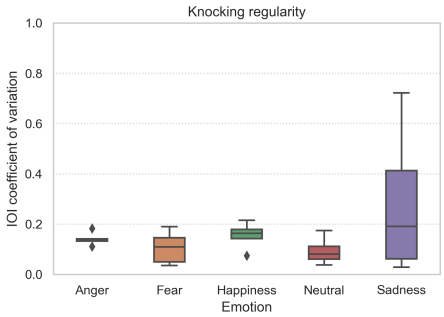

Figure 7. Knocking regularity

4.2.5 *RMS Energy Regression Slope per Action*

- On average, Happiness and Sadness actions decrease in energy.
- On average, Fear and Anger actions increase in energy.
- Anger, Happiness and Fear have a higher variability than Sadness (SD for Anger, Happiness and Fear 2.5 to 5 times higher than the SD for Sadness).

This result relates to the results for sound loudness variability reported in [5], which is high or medium for Anger, Fear and Happiness, and Low for Sadness.

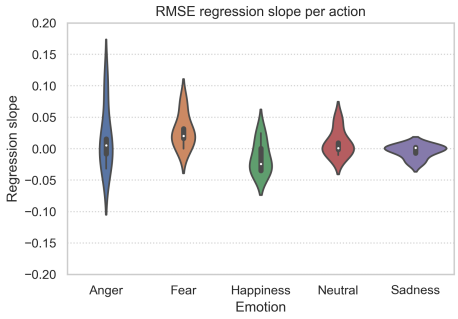

Figure 8. RMS energy regression slope

5. DISCUSSION

The main aim of this study was to verify whether, in knocking actions, intended basic emotions could be perceived through sound alone. The results show that this is possible for Anger, which is the most easily recognisable, and, to a lesser extent, Happiness and Sadness. Results also show that, in this study, Fear tends to be confused with Anger, and Neutral can be confused with Sadness. In regard to the confusion between Fear and Anger, we suggest that this could be due to the scenario we gave to the performers (i.e. communicate a high risk situation), which might have pushed the performance towards a particular type of Fear such as panicked Fear, which perhaps shares characteristics with Anger, rather than, for example, trepidation which could be expressed differently. We note also that the

Neutral state was the most chosen emotion throughout the test (29.8%). We suggest that this might have been used by the listeners as the go to option when the listener was not confident about the choice. The fact, however, that results for most emotions are statistically significant means that listeners were able to make a decision in most cases. From the acoustic analysis, we note that Anger and Fear are characterised by a similar combination of acoustic cues: long knocking actions, a high number of knocks and rate of knocks, regular patterns and crescendo slopes. Happiness is characterised by a high knocking rate, medium duration and number of knocks, regular pattern and decrescendo slope. Sadness presents a medium duration, low number of knocks, low rate, an irregular pattern and decrescendo slope. The Neutral state is characterised by a short duration, low number of knocks and rate, regular pattern and crescendo energy. Furthermore, we looked at similarities between the results of this acoustic analysis and other related studies. There is consistency between the results for Rate, Regularity (for Happiness) and RMSE in this study and results reported in [5] on acoustic cues of emotions in music and vocal expression. This study's results seem also consistent with [23] where "Strong similarities between the use of acoustical features in knocking and music performance were found". We suggest that our findings could further support the hypothesis advanced in [20] that the musical expression of emotions might have a motor origin.

6. FUTURE WORK

Further work is planned to expand the perceptual test to the whole data set; run a larger acoustic analysis study that uses a higher number of samples per emotion class so that patterns of acoustic cues can be identified more precisely; vary performance scenarios and verify to what extent these influence results; compare the emotional variety of knocking action performances of non-performers to that of professional Foley artists; and to verify how other contextual information such as visual cues or ambient sound, contribute to the perception of intended emotions in a knocking action.

Acknowledgments

Part of this work was supported by the NordicSMC Network (Project Number: 86892) and the EPSRC Centre for Doctoral Training in Intelligent Games & Game Intelligence (IGGI) [EP/L015846/1].

7. REFERENCES

[1] P. Ekman, "Basic emotions," *Handbook of cognition and emotion*, vol. 98, no. 45-60, p. 16, 1999.

[2] ——, "An argument for basic emotions," *Cognition & emotion*, vol. 6, no. 3-4, pp. 169–200, 1992.

[3] K. R. Scherer, "Expression of emotion in voice and music," *Journal of voice*, vol. 9, no. 3, pp. 235–248, 1995.

[4] P. N. Juslin and J. Sloboda, *Handbook of music and emotion: Theory, research, applications*. Oxford University Press, 2011.

[5] P. N. Juslin and P. Laukka, "Communication of emotions in vocal expression and music performance: Different channels, same code?" *Psychological bulletin*, vol. 129, no. 5, p. 770, 2003.

[6] D. Västfjäll, "Emotional reactions to sounds without meaning," *Psychology*, vol. 3, no. 8, p. 606, 2012.

[7] M. M. Bradley and P. J. Lang, "Affective reactions to acoustic stimuli," *Psychophysiology*, vol. 37, no. 2, pp. 204–215, 2000.

[8] P. Bergman, D. Västfjäll, A. Tajadura-Jiménez, and E. Asutay, "Auditory-induced emotion mediates perceptual categorization of everyday sounds," *Frontiers in psychology*, vol. 7, p. 1565, 2016.

[9] S. Ntalampiras, "A transfer learning framework for predicting the emotional content of generalized sound events," *The Journal of the Acoustical Society of America*, vol. 141, no. 3, pp. 1694–1701, 2017.

[10] T. L. Bonebright, "Perceptual structure of everyday sounds: A multidimensional scaling approach." Georgia Institute of Technology, 2001.

[11] W. W. Gaver, "How do we hear in the world? explorations in ecological acoustics," *Ecological psychology*, vol. 5, no. 4, pp. 285–313, 1993.

[12] ——, "What in the world do we hear?: An ecological approach to auditory event perception," *Ecological psychology*, vol. 5, no. 1, pp. 1–29, 1993.

[13] S. McAdams, A. Chaigne, and V. Roussarie, "The psychomechanics of simulated sound sources: Material properties of impacted bars," *The Journal of the Acoustical Society of America*, vol. 115, no. 3, pp. 1306–1320, 2004.

[14] O. Houix, G. Lemaitre, N. Misdariis, P. Susini, and I. Urdapilleta, "A lexical analysis of environmental sound categories." *Journal of Experimental Psychology: Applied*, vol. 18, no. 1, p. 52, 2012.

[15] J. Edworthy, S. Loxley, and I. Dennis, "Improving auditory warning design: Relationship between warning sound parameters and perceived urgency," *Human factors*, vol. 33, no. 2, pp. 205–231, 1991.

[16] C. Suied, P. Susini, and S. McAdams, "Evaluating warning sound urgency with reaction times." *Journal of experimental psychology: applied*, vol. 14, no. 3, p. 201, 2008.

[17] X. Li, R. J. Logan, and R. E. Pastore, "Perception of acoustic source characteristics: Walking sounds," *The Journal of the Acoustical Society of America*, vol. 90, no. 6, pp. 3036–3049, 1991.

[18] A. Tajadura-Jiménez, M. Basia, O. Deroy, M. Fairhurst, N. Marquardt, and N. Bianchi-Berthouze, "As light as your footsteps: altering walking sounds to change perceived body weight, emotional state and gait," in *Proceedings of the 33rd annual ACM conference on human factors in computing systems*, 2015, pp. 2943–2952.

[19] R. Nordahl, "Self-induced footsteps sounds in virtual reality: Latency, recognition, quality and presence," in *The 8th Annual International Workshop on Presence, PRESENCE 2005*. Department of Computer Science, Aalborg University, 2005, pp. 353–355.

[20] B. L. Giordano, H. Egermann, and R. Bresin, "The production and perception of emotionally expressive walking sounds: Similarities between musical performance and everyday motor activity," *PloS one*, vol. 9, no. 12, 2014.

[21] K. Amaya, A. Bruderlin, and T. Calvert, "Emotion from motion," in *Graphics interface*, vol. 96. Toronto, Canada, 1996, pp. 222–229.

[22] M. M. Gross, E. A. Crane, and B. L. Fredrickson, "Methodology for assessing bodily expression of emotion," *Journal of Nonverbal Behavior*, vol. 34, no. 4, pp. 223–248, 2010.

[23] R. Vitale and R. Bresin, "Emotional cues in knocking sounds," in *10th International Conference on Music Perception and Cognition, Sapporo, Japan, August 25-29, 2008*, 2008, p. 276.

[24] "SoGoSurvey kernel description," https://www.sogosurvey.com/, accessed: 2020-02-28.

[25] A. Barahona-Ríos and S. Pauletto, "Synthesising Knocking Sound Effects Using Conditional wave-GAN," in *17th Sound and Music Computing Conference*, Online, 2020.