EXPLORING EMOTION PERCEPTION IN SONIC HRI

Adrian B. Latupeirissa
Sound and Music Computing
KTH Royal Institute of Technology
ablat@kth.se

Claudio Panariello
Sound and Music Computing
KTH Royal Institute of Technology
claudiop@kth.se

Roberto Bresin
Sound and Music Computing
KTH Royal Institute of Technology
roberto@kth.se

ABSTRACT

Despite the fact that sounds produced by robots can affect the interaction with humans, sound design is often an overlooked aspect in Human-Robot Interaction (HRI). This paper explores how different sets of sounds designed for expressive robot gestures of a humanoid Pepper robot can influence the perception of emotional intentions. In the pilot study presented in this paper, it has been asked to rate different stimuli in terms of perceived affective states. The stimuli were audio, audio-video and video only and contained either Pepper's original servomotors noises, sawtooth, or more complex designed sounds. The preliminary results show a preference for the use of more complex sounds, thus confirming the necessity of further exploration in sonic HRI.

1. INTRODUCTION

In this paper we present work conducted within the scope of the SONAO project, introduced in [1]. SONAO aims to improve the comprehensibility of robot non-verbal communication (NVC) through an increased clarity of robot expressive gestures and non-verbal sounds. The purpose of the SONAO project is to incorporate movement sonification in Human Robot Interaction (HRI), i.e. to use movement sonification to produce expressive sounds.

Sounds produced by robots can affect the interaction with humans. Nevertheless, sonic HRI is a vastly underexplored direction. The use of sound has been found in previous studies related to the communicaiton of emotional expression of robots. Two common strategies have been the use of recorded acoustic expressions (such as the sound of crying to express sadness) [2] and the use of tone patterns to convey target emotions, such as in the work by Löffler and colleagues [3], in which they utilized *"sine waveform generated by the built-in generator chirp or tone"*. In another work Song and Yamada [4] opted for beep sounds with varying contour slope as presented in [5]. In regard to these studies, it has been noted that *"social interaction between social robots and humans takes place through multimodal interaction"*, and that there is a need to further study the use of robot sound within multimodal interaction [6, p.80].

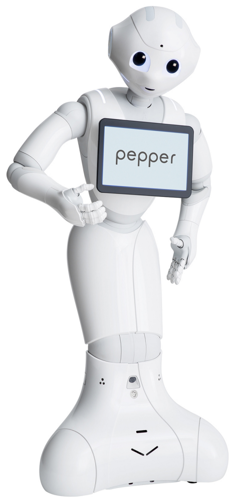


Figure 1: The humanoid robot Pepper

Due to the nature of multimodal interaction, it has been suggested that poorly designed interaction — such as giving a mismatched sound to the appearance or movement of a robot — might introduce undesirable effects in HRI [7].

In the current paper we present the preliminary results of our pilot study focusing on the perception of different sounds set (one more simple, the other more complex) associated with expressive movements of the humanoid robot Pepper, shown in Figure 1. Our expectation is to find a preference towards the use of more refined sounds thus showing the potentiality of better and richer sound design in sonic HRI.

2. METHOD

We carried out an online perceptual rating experiment described below.

2.1 Participants

Participants were recruited among students and colleagues at KTH Royal Institute of Technology and on social networks. A total of 17 participants (6 F, 11 M), age between 23 to 43 years old (average age 30.65) completed the experiment described below. Concerning the participants' musical background, two participants stated to have no experience, two to have little experience, four some experience, three defined themselves semi professional and most of them (six) as expert.

2.2 Material: Video recordings and sound sets

To investigate the perceptual effect of different set of sounds associated with expressive robot movements, we decided to propose the stimuli containing sound (i.e. stimuli and audio-video stimuli) in three different versions: with no additional sounds a part from the original ones produced by the robot itself; with a first set of simple sounds based on sawtooth; with a second set of more complex sounds based on feedback chains.

The sawtooth and the feedback synthesized sounds were overlapped to the original ones. This choice was because on the present state of things there is no possibility to avoid Pepper to produce its own mechanical noises, so having a stimuli without those original sounds would have been unrealistic.

From now on, we will refer to the original sounds as S1; to S1 overlapped to the sawtooth as S2; to S1 overlapped to the feedback synthesized sounds as S3.

The sawtooth and the feedback synthesized sounds were both realized using the SuperCollider programming environment [1].

All the materials (video and sounds, including their spectrograms) are available online [2].

2.2.1 *Robot movement*

Movements of the robot are adapted from the emotional postures defined in [8]. In that study, postures for the humanoid robot NAO have been generated and studied for anger, sadness, and happiness. The five best postures for each emotions identified in that study were incorporated into our Pepper robot. Thanks to the strong similarities of mechanics of the joints between NAO and Pepper, the adaptation faced little problem. The fourth posture that we choose is relaxed, and adapted from Pepper's basic standing posture provided by the robot producer.

The postures are then animated and adjusted following the description of behavior patterns for the associated emotions as defined in [9], such as *"forward body movement"* for anger and *"symmetrical up-down repetitive arm action"* for elated joy. An adjustment was made to the description of sadness (*"both arms at rest and in the pockets"*) with Pepper lowering the arms further down instead.

2.2.2 *Original robot sounds*

The Pepper robot is quite noisy when it is moving and/or performing expressive gestures, due to its size and its 20 servomotors [3]. As already shown in [1], *"the mechanical sounds inherent to the robot NAO's movement appear not to be clearly linked to the emotional scales used in the current study (sad, joyful, frustrated and relaxed)"*. Moreover, *"the sound of a relaxed movement did not necessarily sound very relaxed"*. To investigate how our Pepper robot's mechanical sounds were affecting the perception of the emotions, we decided to have one of the stimuli with these sounds only.

[1] https://supercollider.github.io/
[2] https://kth.box.com/v/robotmovementsounds
[3] http://doc.aldebaran.com/2-4/family/pepper_technical/motors_pep.html

2.2.3 *Sawtooth sounds*

The first sound synthesis method was very simple and based on filtered sawtooth. The `synth` was the same that had been already used to depict the category of pitched sound in [10].

The `synth` had the possibility to be shaped with a parametric envelope and to be granularized using the `GrainIn` unit generator. The output was then sent into a reverberation module.

Prior studies have already proven the efficiency of these basics sounds in sonic HRI, as the BEST (Bremen Emotional Sound Toolkit) database shows [11].

To create appropriate sounds for the robot's gestures, temporal marks were inserted into the video in correspondence of the arms' movement. The timestamps were then used to create a score on SuperCollider's client side using the `Task`, that is a pauseable process. An array was initialized with the timestamps and it was used to schedule all the sound events, specifying the parameter variations in the Synth in terms of pitch, envelope and filtering. The choice of the parameters was informed by the characteristic values of musical variables used in the communication of emotions in music performance as reported in previous research [12]. For the generation of sounds communicating excitement we used a higher overall amplitude compared to the other emotions, and their pitch range was wider and centered on medium-high pitches (from 65 to 89 MIDI notes, two octaves span). On the other hand, the sounds for communicating anger and the sadness had narrower ranges centered on medium-low pitches (respectively from 50 to 64 MIDI notes, and from 48 to 54 MIDI notes). The relaxed sounds had even a smaller range, oscillating between 66 and 68 MIDI notes. The cutoff frequency was chosen in order to go from a muffled quality for the relaxed sounds on one end to a brighter one for the angry sound on the opposite end. The duration of the envelope's attack went from 0.01s for the excited sound, to 0.05s for the angry sound, to 0.1s for the relaxed and sad ones.

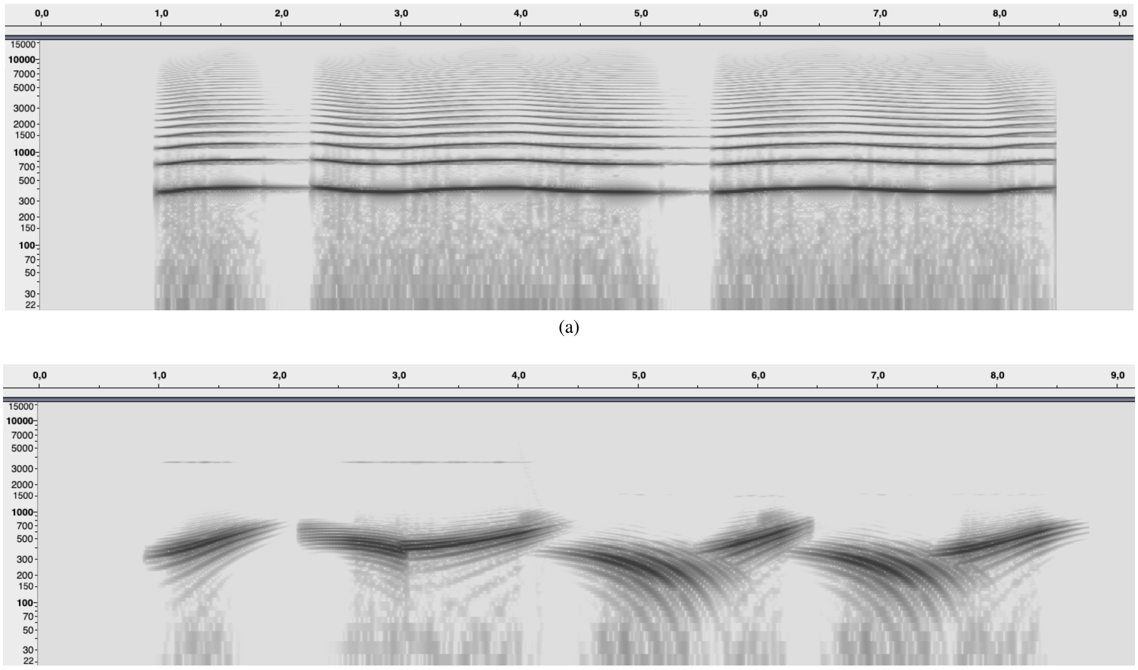
The results were recorded and finally superimposed on the videos. The SC patch is summarized in the code presented in Listing 1. The spectrogram of one of the sawtooth sounds thus synthesised, namely the relaxed one, is shown in Figure 2a.

The SuperCollider patch is available online [4].

Listing 1: SuperCollider patch structure.

```
//Definition of Synths
SynthDef(\pitch, {...}).send(s);
SynthDef(\rev, {...}).send(s);
//
(
//Score
~timestamp = [...];
t = Task({
        s.record(path +/+ "result.aiff");
        r = Synth(\rev);
        ... //Sound events//...
}).start;
)
```

[4] https://kth.box.com/v/robotmovementsounds



(a)

(b)

Figure 2: Spectrograms of the synthesised relaxed sounds, one synthesised with sawtooths (a) and the other one with feedback chains (b). On the x-axis there is time (s) and on the y-axis frequency (Hz).

2.2.4 *Synthesized sounds with feedback chains*

The second sound synthesis was more complex and refined, and based on feedback chains. This synthesis was chosen because it had been previously developed with clear esthetical aims and not in order to fulfill a specific functionality, thus meaning its intrinsic sound quality and timbrical search were in foreground.

The feedback chain `feedback_synth` is excited by an impulse of a non-band-limited pulse oscillator. The impulse is written on an internal bus `LocalOut` and it is filtered by a bandpass filter `BPF` which cutoff frequency can dynamically move between two extremes. Therefore, spectral processes are applied to the signal: shifting of the phases and randomization of bins' order. The window size chosen for the FFT analysis affects the rhythmicity of the result. Then the output returns to itself multiplied by a feedback factor `fb`. An envelope follower on the resulting sound is used as a negative feedback control signal in order to prevent the saturation.

To create the sounds, the same score developed for the previous case was used (subsection 2.2.3), adjusting the parameters according to the needs of the `feedback_synth`. For the sad sound, a 50% randomization of the bins' order was chosen and the extremes of the bandpass filter were extremely wide (from 100 to 10000 Hz), with a window size of 4096 samples, thus resulting in a slow moving sound with an inharmonic spectrum. For the excited sound, two `feedback_synth` were called at once, with 30% randomization and a wide bandpass filter range (from 200 to 10000 Hz), with a window size of 512 samples, thus resulting in a fast moving sound with a more compact spectrum. For the relaxed sound, again two `feedback_synth` were called at once, with 5% randomization and a very narrow bandpass filter range (from 66 and 68 MIDI notes, as in the relaxed sawtooth case), with a window size of 1024 samples, thus resulting in a medium-fast moving sound with a very compact spectrum. For the angry sound, 10 `feedback_synth` were called at once, with 50% randomization and a very lower bandpass filter range (from 100 to 500 Hz), and with window size of 1024 samples, thus resulting in a medium-fast moving inharmonic spectrum with great energy in the low frequencies.

An important difference with the *Synthesis with sawtooth* case is that the feedback sounds were not always the same, by their very nature. Therefore, since the overall process of sound creation in this second stage more oriented towards sound design, more layers were created for each of the four affective states, and then edited in a multitrack environment, in order to achieve a richer and more structured result.

The entire process and the `feedback_synth` are summarized in the code presented in Listing 2. The spectrogram of one of the sounds thus synthesised, namely the relaxed one, is shown in Figure 2b.

The SuperCollider patch is available online [5] .

[5] https://kth.box.com/v/robotmovementsounds

Listing 2: Feedback patch structure.

```
//Definition of Synths
(
~busenv = Bus.audio(s, 1);
SynthDef(\impulse, {...}).send(s);

SynthDef(\feedback_synth, {arg ...;
        var ...;

        feedback = LocalIn.ar(1)*fb;
        in = In.ar(~busenv, 1);
        out = LeakDC.ar(feedback + in);
        output1 = BPF.ar(out, ...); //
            bandpass filter
        chain = FFT(LocalBuf(windowsize, 1)
            , ...);
... // spectral transformations //...
        control = ...//control function

        pan = Pan2.ar(Limiter.ar(output1*
            control), 0, Lag2.kr(level,
            lagLevel));
        Out.ar(0, pan);
}).send(s);
)

// sound generation
(
2.do({Synth(\feedback_synth, [...]);})
Synth(\impulse);
)
```

2.3 Procedure

We carried out an online perceptual rating experiment [6]. The survey was shared with students and colleagues at KTH Royal Institute of Technology and on social networks. Participants were allowed to re-distribute the survey to anyone they believed would be interested in participating in the study.

Participants were asked to rate stimuli on four semantic opposites corresponding to four emotions (sadness, excitement, anger, relaxation), and on five semantic opposites for describing other qualities of the robot behaviour (pleasantness, typicality, efficiency, likeability, trust). The four emotional scales were chosen in order to cover each of the four quadrants of the circumplex space of emotions [13], and they were also resulted to be the most successful ones to be recognized in previous robot research [8].

The five qualities for describing the robot behaviour were inspired by a previous study in which researchers identified properties relevant for describing vacuum cleaner sounds [14] and by studies on trustworthiness in HRI [15, 16].

The stimuli presented in the survey were: audio, audio-video and video only, with four stimuli for each category corresponding to each of the emotional intentions. Moreover, we had three different set of sounds S1, S2, S3 (as explained in Subsection 2.2), thus making a total of 28 stimuli. The stimuli were audio-video recordings of expressive gestures performed by a Pepper robot, with or without one

[6] The entire survey can be accessed at https://www.surveygizmo.com/s3/5491700/Robot-Movement-Sound.

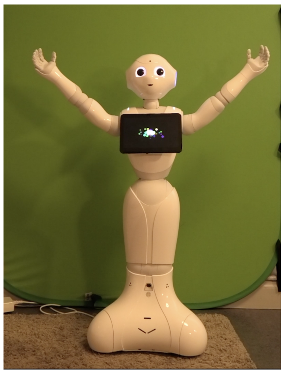


Figure 3: Screenshot of video stimulus of the Pepper robot performing an *"excited"* gesture.

of the three different sets of sounds. A screenshot of a video stimulus is shown in Figure 3.

Presentation order was randomized for each participant. The participants could rate the stimuli on nine scales corresponding to the semantic opposites listed above. The four scales regarding the emotions were going from "not at all" (0) to "very much" (5.0) with step size of 0.1 (e.g. from "not at all sad" to "very much sad"; the other five scales varied from one perceptual category (e.g. "unpleasant") to its opposite ("pleasant"), with the same range and step size as for the emotion scales.

3. RESULTS

Descriptive statistics for the four emotion ratings (angry, excited, relaxed and sad) are shown in Figures 4a-4d. Descriptive statistics for the five behaviour ratings (pleasantness, typicality, efficiency, trust, likeability) are shown in Figures 5a-5e. All the graphs show median values. We will refer to the video only with V-stimuli; to the sounds only with the already mentioned S1, S2, S3; to the three combination audio-video with VS1, VS2 and VS3. Full statistical analysis report is available online [7].

3.1 Emotion

For emotion ratings, two-way repeated measures ANOVA analysis was conducted to observe the significance of sound (four levels: no sound, S1, S2, and S3), video (two levels: no video and V), and the interaction between the two. Zero ratings were added for the combination of no video and no sound. While participants responded with ratings for all emotional perceptions in each of the stimuli, the ANOVA analysis was conducted only on the ratings of the correct emotion since we are only interested to see the significance of the stimuli.

3.1.1 Anger

For the angry stimuli we found that V was recognized as excited. The original sound of the robot (S1) was rated as both excited and angry, while S2 and S3 were not rated

[7] https://kth.box.com/v/robotmovementsounds

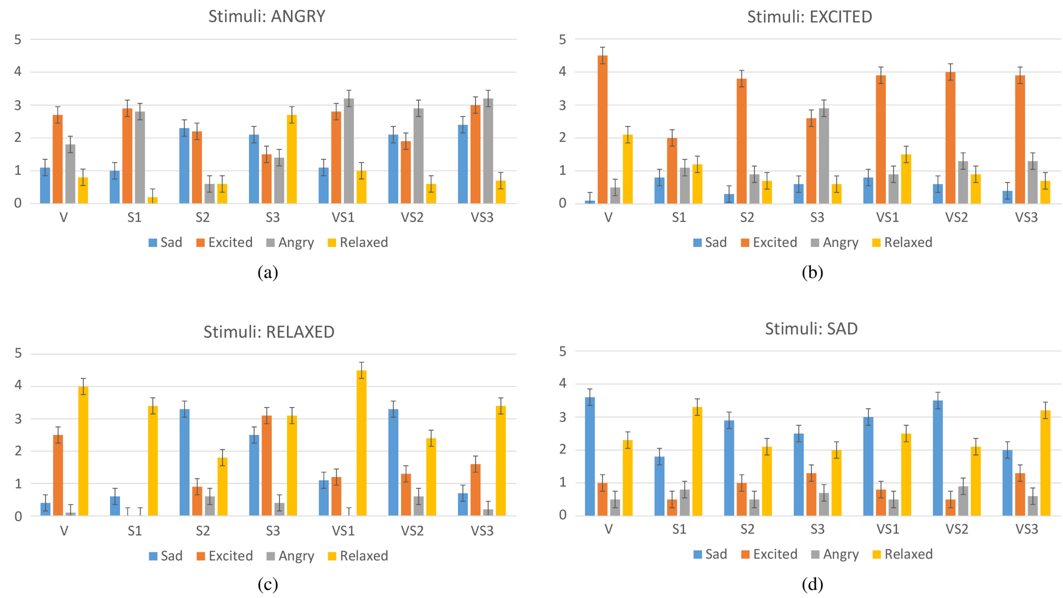


Figure 4: Median perceptual ratings for respective stimuli category with 5% error bars.

as angry and were confused with other emotions. Interestingly, the stimuli combining the angry video with sounds were rated as angry. This could be an audio-visual crossmodal effect. ANOVA analysis shows that the effect of the interaction sound * video is significant, $F(3, 48) = 3.87$, $p = 0.015$.

3.1.2 *Excitement*

For the excited stimuli we found that V and S2 were clearly identified as excited. S1 was rated as excited but less than S2 and V, while S3 was confused with angry. Nevertheless VS1, VS2 and VS3 were all rated as excited. ANOVA analysis shows that the effect of the interaction sound * video is significant, $F(3, 48) = 24.74$, $p = 0.000$.

3.1.3 *Relaxation*

For the relaxed stimuli we found that V was clearly rated as mostly relaxed and, to a lesser extent, as excited. S1 was clearly rated as relaxed, while both S2 and S3 were confused with other emotions. The stimuli combining sound and video in VS1 and VS3 were rated clearly as relaxed. However, VS2 was rated mostly as sad. ANOVA analysis shows that the effect of the interaction sound * video is significant, $F(3, 48) = 14.29$, $p = 0.000$.

3.1.4 *Sadness*

For the sad stimuli we found that V was classified as sad and, to a lesser extent, relaxed. S2 and S3 were both rated as sad but S1 was identified as relaxed. VS1 and VS2 were both rated as sad and VS3 was rated more relaxed than sad. ANOVA analysis shows that the effect of the interaction sound * video is significant, $F(3, 48) = 13.48$, $p = 0.000$.

3.2 Behaviour

In this section, given the large amount of data which could be analyzed and the limited space available for this paper, we limit our report of results for the five qualities describing the robot behaviour only to the stimuli combining sound with video (VS1, VS2, VS3). Two-way repeated measures ANOVA analysis was conducted to observe the significance of emotion (four levels: anger, excitement, relaxation, and sadness), sound (three levels: VS1, VS2, VS3), and the interaction between the two.

3.2.1 *Pleasantness*

We found that all the angry stimuli were rated as less pleasant than all the other stimuli, regardless S1, S2 or S3. While the effect of the interaction emotion * sound is not significant, $F(6, 96) = 2.115$, $p = 0.058$, individually the effect of the emotion displayed by the robot [$F(3, 48) = 10.585$, $p = 0.000$] and the sound [$F(2, 32) = 3.43$, $p = 0.045$] are significant.

3.2.2 *Likeability*

We observed that VS2 stimuli were in average the less liked by the participants. The VS3 relaxed stimulus was rated as the most likeable one, followed in order by the VS2 sad one and by the VS1 relaxed one. ANOVA analysis shows that the effect of the interaction emotion * sound is significant, $F(6, 96) = 3.23$, $p = 0.006$.

3.2.3 *Efficiency*

We found that all sounds were rated almost equally in how efficient the sounds were to communicate an emotion. However, slightly higher ratings were given toward VS1 in angry and VS3 in sad. ANOVA analysis shows that the effect

Figure 5: Stimuli combining sound with video: median perceptual ratings of the for respective behaviour quality with 5% error bars.

of the interaction emotion * sound is significant (corrected with Greenhouse-Geisser, $\epsilon = 0.601$), F(3.61, 57.735) = 2.79, p = 0.039.

3.2.4 *Typicality*

We found that while both VS1 and VS2 was rated more typical in in showing the emotion with high arousal (excited and angry), VS1 received higher rating. VS3, on the other hand, rated slightly higher on the relaxed emotion and lower on angry. ANOVA analysis shows that the effect of the interaction emotion * sound is significant (corrected with Greenhouse-Geisser, $\epsilon = 0.618$), F(3.71, 59.375) = 4.07, p = 0.009.

3.2.5 *Trustworthiness*

We found that the angry stimuli were rated as less trustworthy than all the other stimuli, regardless S1, S2 or S3. The sad and excited stimuli were also rated as the most trustworthy for VS1 and VS2, although in VS3 the most trustworthy were sad and relaxed. ANOVA analysis shows that the effect of the interaction emotion * sound is significant (corrected with Greenhouse-Geisser, $\epsilon = 0.601$), F(3.6, 57.74) = 2.85, p = 0.036.

4. DISCUSSION

Our preliminary results show that the video only stimuli for excited, relaxed, and sad emotions appear to be successful in communicating the respective affective states. On the other hand, the anger video stimulus is confused with the excited one. Interestingly, when the audio is added to it (either the recorded original sound S1, the sawtooth overlapped with S1 – namely S2 –, or the synthesized sound overlapped with S1 – namely S3) the perception changed and participants rated it mostly as angry (see Figure 4a). The reason why the perception changes might be explained considering that excitement and anger have a similar arousal but opposite valence [13], and the sound is helping in compensating it.

The excited stimuli are perceived as such both with and without sound (see Figure 4b). This can be explained taking into consideration the high arousal and valence of that excited gesture, probably hard to be confused with other ones.

The ratings for the relaxed and sad V-stimuli show that participants could clearly identify those intended emotions (Figures 4c and 4d). Very similar results were found also when S1 was added. However, the addition of S2 or S3 to the video changed its perception: the relaxed stimuli with S2 was perceived mostly as sad; the sad stimuli with S3 was mostly rated as relaxed.

Considering the S2 ratings for its angry, relaxed and sad versions (see Figures 4a, 4d and 4c), it seems that S2 is perceived in all three cases predominantly as sad. This suggests that the quality of S2 could in general carry both low arousal and low valence information, thus accordingly affecting the perception of the expressive gestures shown in the video.

Looking at the five robot behaviour qualities, we see that the pleasantness score (Figure 5a) of VS2 is, except for the excited affective state, always lower than the cases with VS1 and/or VS3. This is not only coherent with what has already been said regarding some inner sad quality of VS2 (if a sound is perceived as sad, it will be most likely rated as unpleasant), but it is also suggesting to look for better designed sounds. Also the likeability of VS2 follows the same pattern and, to a lesser extent, its efficiency (see Figures 5b and 5c).

The typicality quality, shown in Figure 5d has been found to be higher for VS1. Possibly, the mechanical servomotors noises were regarded as the only ones to be expected while interacting with the Pepper robot. Another possible explanation could be that participants, probably the ones with most musical expertise, might have spotted the superposition of VS2 or VS3 to VS1.

According to the data we have at the present moment, it seems that the trust quality is not significantly affected by the sounds (see Figure 5e).

5. CONCLUSIONS AND FUTURE WORK

In this paper we have presented the preliminary results of our study within the SONAO project focusing on perceptual ratings of different sets of sounds communicating expressive movements of the Pepper robot also in combination with videos of the same movements. Although more data collection and analysis must be carried out, our preliminary results suggest that:

- S3 were generally liked the most and regarded as more pleasant by test participants. Theses sounds are characterized not only by smoother attacks but especially by richer spectra with partials distributed in a more complex way compared to the ones generated with sawtooths, thus resulting in more appealing timbres thanks to spectral fusion. This suggests to investigate the use of complex synthesis methods producing richer robot sounds than it has been done in the past (see for example [3, 4, 11]). This could produce different results in HRI research and improve interaction with robots;

- Results suggest that S2 can be more suitable for representing sad affective states, and can help in communicating excitement as well;

- The excited stimuli were the less likely to be misunderstood independently from the sound used, probably because of the prominence of the original sound of the motors (S1) always embedded in the synthesized sounds (S2 and S3). This suggests that for achieving a clearer communication of robot expressions characterized by a low activity, one should make use of either masking sounds or strategically placed loudspeakers depending on the mechanical activity of the robot.

We are aware that, given the limited number of subjects (17) and their composition, the results of this study can not be generalized, but they show possible tendencies. Therefore, the study presented in this paper will be followed up by more detailed evaluation of the stimuli and we are also planning to test different synthesis methods in the near future with a larger set of participants.

Acknowledgments

This project was funded by Grant 2017-03979 from the Swedish Research Council and by NordForsk's Nordic University Hub "Nordic Sound and Music Computing Network - NordicSMC", project number 86892.

6. REFERENCES

[1] E. Frid, R. Bresin, and S. Alexanderson, "Perception of Mechanical Sounds Inherent to Expressive Gestures of a NAO Robot-Implications for Movement Sonification of Humanoids," in *Proceedings of the Sound and Music Computing Conference*, 2018.

[2] M. Häring, N. Bee, and E. André, "Creation and evaluation of emotion expression with body movement, sound and eye color for humanoid robots," in *2011 RO-MAN*, July 2011, pp. 204–209.

[3] D. Löffler, N. Schmidt, and R. Tscharn, "Multimodal expression of artificial emotion in social robots using color, motion and sound," in *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2018, pp. 334–343.

[4] S. Song and S. Yamada, "Expressing emotions through color, sound, and vibration with an appearance-constrained social robot," in *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ser. HRI '17. New York, NY, USA: ACM, 2017, pp. 2–11. [Online]. Available: http://doi.acm.org/10.1145/2909824.3020239

[5] T. Komatsu, "Toward making humans empathize with artificial agents by means of subtle expressions," in *International Conference on Affective Computing and Intelligent Interaction*. Springer, 2005, pp. 458–465.

[6] S. Yilmazyildiz, R. Read, T. Belpeame, and W. Verhelst, "Review of semantic-free utterances in social human–robot interaction," *International Journal of Human-Computer Interaction*, vol. 32, no. 1, pp. 63–85, 2016.

[7] C. McGinn and I. Torre, "Can you tell the robot by the voice? an exploratory study on the role of voice in the perception of robots," in *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2019, pp. 211–221.

[8] M. Erden, "Emotional postures for the humanoid-robot Nao," *International Journal of Social Robotics*, vol. 5, no. 4, pp. 441–456, 2013.

[9] N. Dael, M. Mortillaro, and K. R. Scherer, "Emotion expression in body action and posture." *Emotion*, vol. 12, no. 5, p. 1085, 2012.

[10] C. Panariello, S. Sköld, E. Frid, and R. Bresin, "From vocal sketching to sound models by means of a sound-based musical transcription system," 2019, pp. 1–7.

[11] A. Kappas, D. Küster, P. Dente, and C. Basedow, "Simply the B.E.S.T.! Creation and Validation of the Bremen Emotional Sounds Toolkit," *International Convention of Psychological Science*, 2015.

[12] R. Bresin and A. Friberg, "Emotion rendering in music: Range and characteristic values of seven musical variables," *Cortex*, vol. 47, no. 9, pp. 1068–1081, 2011.

[13] J. A. Russell, "A circumplex model of affect." *Journal of personality and social psychology*, vol. 39, no. 6, p. 1161, 1980.

[14] L. Kynčl and O. Jiříček, "Psychoacoustic product sound quality evaluation," in *17th International Congress on acoustics*, 2001.

[15] P. A. Hancock, D. R. Billings, K. E. Schaefer, J. Y. C. Chen, E. J. de Visser, and R. Parasuraman, "A meta-analysis of factors affecting trust in human-robot interaction," *Human Factors*, vol. 53, no. 5, pp. 517–527, 2011, pMID: 22046724. [Online]. Available: https://doi.org/10.1177/0018720811417254

[16] K. E. Schaefer, J. Y. C. Chen, J. L. Szalma, and P. A. Hancock, "A meta-analysis of factors influencing the development of trust in automation: Implications for understanding autonomy in future systems," *Human Factors*, vol. 58, no. 3, pp. 377–400, 2016, pMID: 27005902. [Online]. Available: https://doi.org/10.1177/0018720816634228