The natural language of robots

Linda Cnattingius
lindacn@kth.se

Martin Linder Nilsson
hmni@kth.se

Jesper Englund
jeengl@kth.se

Johannes Loor
loor@kth.se

Royal Institute of Technology, School of Electrical Engineering and Computer Science,
The Department of Media Technology and Interaction Design

ABSTRACT

The natural language of robots is believed to be grounded in popular culture and therefore establishing a common ground might be a difficult task. But the general understanding is that the fidelity in the appearance of a humanoid robot ought to correspond to the fidelity of the sounds it emits. However, somewhere along the way, this notion supposedly intersects with *the uncanny valley*, the concept that humanoid objects almost resembling actual humans provoke unpleasant feelings in the observer. In this study, we explore this intersection by letting the humanoid robot Pepper convey four different emotions through movements accompanied by two different sets of non-verbal sounds. One of the sets was composed of recorded human sounds and the other was composed of sinusoids. These two sets of sounds were assigned to different test groups. The participants evaluated their interactions with the robot, and by comparing the evaluations from the different test groups, we hope to shed some light on what robots should sound like when conveying emotions.

1. INTRODUCTION

Expectations in robot behaviour may originate largely from films and popular culture in general [4]. How emotions through behaviours are most appropriately and effectively conveyed is therefore a subject of controversy and under constant development and research. As of now, there is no one true answer to this question. We wanted to investigate the matter further, and focus on conveying emotions through non-verbal audio feedback. In particular, to compare sinusoid feedback with humanlike feedback, and see what we perceive as most natural, comprehensible and appealing. This was investigated by using Pepper, a humanoid robot. Pepper is a robot optimised for human interaction [9] and has an open and programmable platform where one can customize the movements, sounds and interactions the robot can generate.

In light of the expectations we have on robot audio feedback, the answer is not clear and is therefore an interesting topic to examine. The topic is especially relevant in these times of technological advancements, where interactions between humans and robots or AI is rapidly becoming more mundane and a frequent phenomenon in people's everyday life.

2. RELATED WORK

Häring et al. [3] evaluated emotion expression with body movements, sound and eye color. The evaluation was based on the Pleasure-Arousal-Dominance (PAD) model and resulted in a number of different more or less successful expressions of joy, sadness, fear and anger. Erden [1] tested different postures of a NAO robot and evaluated what postures were most associated with happiness, sadness and anger. The movements and the sinusoid sounds used in the present study were heavily influenced by some of the more successful expressions from these two studies. Song and Yamada conducted a study focusing on the expression of emotion in appearance constrained robots [8]. Modalities explored were color, sound and vibration, of which only sound was used in the creation of sinusoids in the present study. Among other points of interest in the study the authors concluded that negative emotions is easier to convey than positive emotions and that sadness is preferably expressed through a falling sound. In a study [5], Lima et al has created a "corpus of nonverbal vocalizations for research on emotion processing" where voice actors were recorded expressing eight different emotions in several variations. The library of sounds was rated and evaluated according to the emotions they conveyed as well as valence, arousal and authenticity. All nonverbal human sounds used in the present study were chosen from this library. Research by Latupeirissa et al [4] suggests that the physical appearance of robots in films, is often connected to their sonic representation. By analyzing several pop-cultural depictions of robots they conclude that robot voices show significant differences from human voices in regard to frequency representation. This was taken into account in the process of choosing and creating the sounds for the present study. Frid et al [2] conducted a study exploring whether the mechanical sounds inherent in robots could be carriers of emotion in themselves. Results show that mechanical sounds prove to be poor conveyors of emotion in general, and that they are more suitable expressing arousal than valence.

3. AIM AND HYPOTHESIS

The aim of this research is to provide a better understanding of how human qualities are perceived in robots, with a focus on sonic representation in particular. This could potentially be of use in future audio design of robots or in further research on the subject. In the process of understanding the role that sonic representation has in making us recognize a mechanical entity as a robot, we arrived at the following research questions:

- What is the natural language of robots as perceived by us humans?
- To what extent has popular depictions of robots altered our sonic expectations of them?

- Do human sounds have a tendency to produce an uncanny feeling in us when coming from a robot?

Our hypothesis is that the influence of pop-culture has had a real impact on how we expect robots to sound. However, we expect that the human emotion sounds may be easier to interpret than the sinusoids by test subjects. As for the question about whether a human voice in robots will induce an unnatural feeling, we believe that the study will show that this is the case.

4. METHOD

The study was conducted with the help of 18 test subjects on october 9, 2019. The participants were between 22 and 27 years old, most of them graduate engineering students in media technology at KTH. Two subgroups were formed, Group A and Group B, each consisting of 9 people. Tests were carried out using a Pepper robot programmed to imitate human emotion through movement and sound which was controlled with the Choregraphe software[10]. Two sets of sounds were created for the experiment, one set with recordings of human non-verbal expressions, and the other consisting of sinusoid sounds at different pitches and intonations.

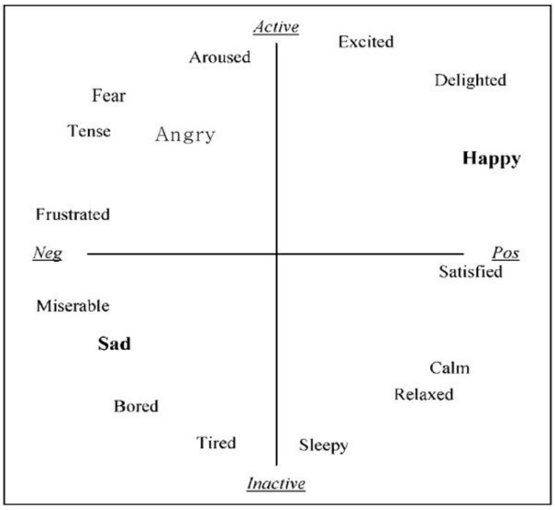

Figure 1. The circumplex model of affect. Valence is displayed on the horizontal axis and arousal on the vertical axis.

4.1. Sound

The set of human sounds were chosen from the "corpus of nonverbal vocalizations" created by Lima et al [5] containing human reaction sounds expressing different emotions. All sounds were made by the same male voice actor, raised in pitch by three semitones to accommodate for Pepper's small stature. The sinusoid sounds were all made by a single sine wave oscillator using the Pure Data software [11], and were created to mimic these emotions using variations in pitch and intonation. Inspiration for the design of the sinusoids was taken from previous research working with robot sounds [3,8], as well as own creations and alterations to make the robot's movement and sounds fit together in a natural way.

Four different emotions were chosen corresponding to four diametrical points in the circumplex model of affect[6], shown in Figure 1. These emotions were:

- Sadness - Corresponding to low valence and low arousal.
- Fear - Corresponding to low valence and high arousal.
- Relief - Corresponding to high valence and low arousal.
- Joy - Corresponding to high valence and high arousal.

The human sound expressing sadness was a sobbing sound while the sinusoid was represented by a sine wave slowly falling in pitch. Fear was represented by a gasp by the voice actor and, in the case of the sinusoids, a note quickly increasing in pitch landing in a vibrating high pitch note. The relief emotion was expressed through a sigh of relief by the voice actor, i.e an inhaling sound followed by a sigh, while the sinusoid mimicked this sound with a fast increase in pitch followed by a slower decreasing pitch note. Joy was expressed by laughing for the human set of sounds and, for the sinusoid set, two consecutive increasing pitch notes.

4.2. Movement

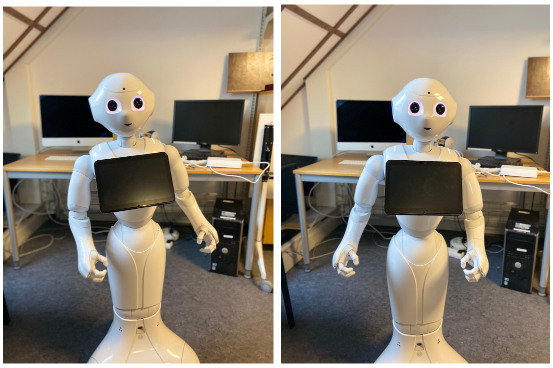

Figure 2. Pepper expressing joy.

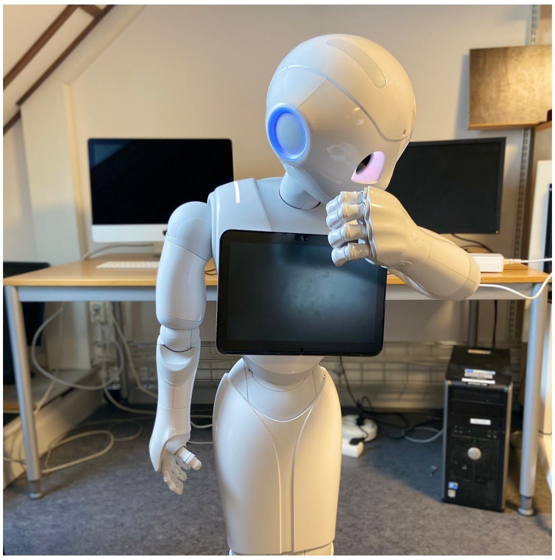

Figure 3. Pepper expressing sadness.

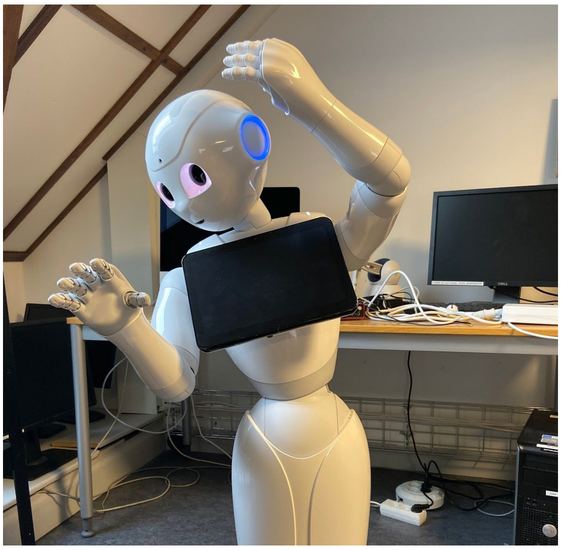

Figure 4. Pepper expressing fear.

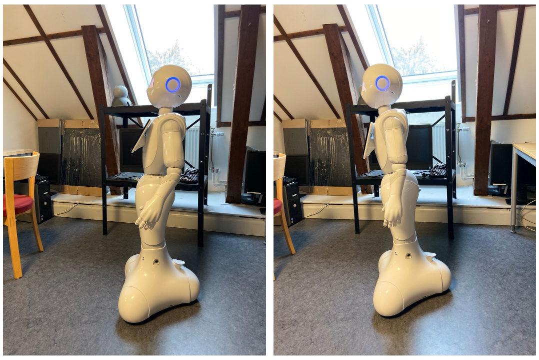

Figure 5. Pepper expressing relief.

The movements for the four expressions were all created in Choregraphe, with the exception of the movement expressing joy, which was a preset movement found in the program. In this movement, Pepper's arms moved from side to side with hands at waist height while the head was bobbing back and forth, see figure 2. Although this motion did not exactly resemble any of the movements recommended in the related work[1,3], it was assessed to fit the bill of a joyful expression. In the movement created for sadness, shown in figure 3, the robot was crouching its upper body and covering its face in its armpit, as if it was crying.

As seen in figure 4, fear was expressed by having the robot twist its body, lean back and cover its face with one of its arms. These two movements corresponded partially to the examples that were discovered in the related work[1,3]. But the movements also had to be customized for the sonic expressions, which was the main cause for deviation from the examples found in the literature. Another factor here is that a NAO robot was used in these examples, whereas we used the Pepper robot, and although their limb structure is similar, it is not the same. For the relief movement, no previous examples were found. But the motion created was rather simple and candid. To mimic a sigh of relief, the robot leaned back slightly and then slowly leaned forward again to its original position, see figure 5.

4.3. Procedure

Before the experiment test subjects got a brief explanation of how the test was going to work. They were instructed to sit down opposite the Pepper robot, read stories to it slowly and in an expressive way, and later evaluate the experience. Four different stories (see appendix) were read, each representing one of the emotions in the study and each ending with a clear emotional cue, at which Pepper executed a pre-programmed series of movements and sounds conveying an emotional response to the story. These responses were initiated at the cues by test conductors through a software interface from an adjacent room. Both subgroups got the same response experience regarding Pepper's movements, the difference being the auditory feedback as group A got the human feedback sounds and group B got the sinusoids. After all four stories had been read the participant was instructed to answer a short survey evaluating their experience with the following questions:

1. How old are you?
2. What gender do you identify as?
3. How well did you think Pepper expressed joy when reacting to story 1?
4. How well did you think Pepper expressed sadness when reacting to story 2?
5. How well did you think Pepper expressed relief when reacting to story 3?
6. How well did you think Pepper expressed fear when reacting to story 4?
7. When reading the stories, how genuine did you find Pepper's overall reactions to be?
8. How natural did you perceive Pepper's speaking voice to be? Did Pepper's voice/sound meet your expectations of what a robot should sound like? Please explain why
9. How pleasant did you find your interaction with Pepper?

The age and gender questions were answered with a numeric answer and one of the options “Male”, “Female”, “Other” or “Prefer not to say”, respectively. For questions 3 to 6 the answers were given on a scale from 1 to 6 where the ends of the scale corresponded to “Not well” and “Well”. The fact that the scale had an equal number of choices was decided on to avoid the temptation of only choosing the median answer. The options for questions 7 to 9 were also given on the same numbered scale but with “Not genuine” to “Genuine”, “Not natural” to “Natural” and “Not pleasant” to “Pleasant” as the ends of the spectrum. For question number 8, a free text answer was added in addition to the scale to get a better understanding of what the subjects perceived as natural, concerning the speaking voice of robots.

5. RESULTS

Of the 18 participants, 10 chose the gender female, 8 chose male and all were between 22 and 27 years old. When looking at the data divided into groups of male and female, no distinctions could be found. Therefore the results below do not take gender or age into account and the test group is seen as homogeneous.

5.1. Expressiveness of Emotions (question 3-6)

When analyzing the answers to the questions regarding how well Pepper expressed different emotions (joy, sadness, relief and fear), we see that the human sounds got a higher result on average compared to the sinusoids. This is true for all emotions except fear were the sinusoids got a slightly higher average value, see figure 6. As shown in table 1, the retrieved average values for the human sounds (in the order of joy, sadness, relief and fear) are 4.56, 4.00, 4.89 and 4.44. For the sinusoids the values are 4.44, 3.78, 4.00 and 4.56. The difference between the two show an indication that the emotion in the human sounds often were easier to perceive, but this could not be statistically proven (p > 0.05).

Table 1. Collection of mean scores for every question of the evaluation.

	Human avg	Human Stdev	Sinusoid Avg	Sinusoid Stdev
Joy	4.56	1.01	4.44	1.24
Sadness	4.00	1.32	3.78	1.48
Relief	4.89	1.17	4.00	1.50
Fear	4.44	1.24	4.56	1.88
Genuine	3.67	1.00	3.89	0.78
Natural	3.78	1.30	5.11	0.78
Pleasant	4.33	1.22	5.11	0.93

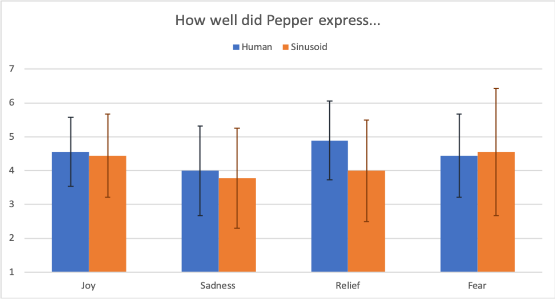

Figure 6. Comparison of the mean scores for how well the emotions were expressed between humanlike sounds and sinusoids. The error bars represent the standard deviations.

5.2. Authenticity (question 7)

The answers to the question about how genuine the test subjects found Pepper’s reactions, showed a higher average for the sinusoids (3.89) compared to the human sounds (3.67). Although not statistically significant ($p > 0.05$), this result is of interest because of how it contradicts the results regarding the emotions.

5.3. Sonic expectations (question 8)

As seen in figure 7, when asked about how natural the subjects perceived Pepper's voice to be, they found the sinusoids to be the most natural. They received an average of 5.11 where as the human sounds average was 3.78. The difference between the two are statistically significant ($p = 0.018$).

When answering this question the participants had the possibility to express their impression of how natural Pepper's sounds were with their own words. The participants explained their sonic expectations of the robot and whether Pepper's voice met these expectations.

5.3.1. Opinions of group A

For the participants who received the human sounds during the test, group A, two main themes could be retrieved from the free text answers.

Quality

An apparent theme from Group A was the expression about the quality of the human sounds. Participants described the human sounds as mechanic, and low quality.

Impression

The participants of group A expressed contrasting thoughts about the overall impression of Pepper. While some participants described the impression of Pepper as creepy or scary others were happily surprised that it was not as creepy as they suspected. Some expressed that the creepiness was because of how the audio feedback sounded as recordings of humans. A participant expressed:

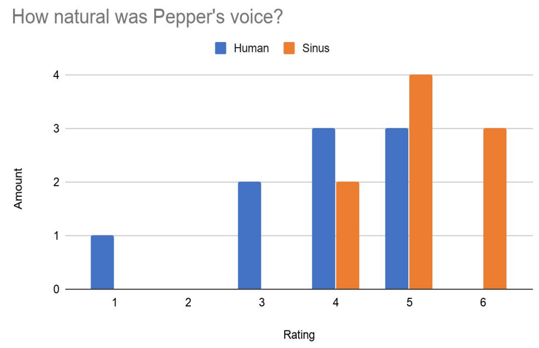

Figure 7. Comparison of the amount of scores rated on a 6-step scale between humanlike sounds and sinusoids.

"It felt a bit creepy that the voice almost sounds like a recorded human. I would have preferred something more artificial, I think"

However, some participants also expressed the sounds as still having "robot-like" characteristics while simultaneously sounding similar to a human.

5.3.2. Opinions of group B

The following will discuss the opinions and impressions of the group who listened to sinusoids during the test.

Quality

Regarding the quality of the sinusoid sounds a few participants described them as "pitchy" and "staccato" but also as kind and more uniform sounding than expected.

Expectations

Group B specifically expressed more about whether their expectations were met compared to group A. The majority of the participants in group B expressed clearly that Pepper's sounds met with their expectations of what a robot sounds like. A few participants expressed the following:

"It met my expectations of a kind robot, it sounded pretty human but a bit staccato."

"I expect robots to make sounds like Pepper, and not speak, so that was good."

"I wasn't expecting the robot to make 'human' sounding noises, that would probably have felt less genuine or natural than a robot-y voice! Considering this the sounds Pepper made felt fitting."

The participants also described the sounds as similar to how pop culture illustrates robot sounds. One participant expressed:

"Pepper pretty much sounds like any robot depicted in today's popular media."

Another participant expressed similarities in sound characteristics with robots from the movie "Star Wars".

5.4. Congeniality (question 9)

The question about how pleasant the interaction with Pepper felt, was asked to get an indication if and how the "uncanny valley" phenomenon was a factor present in the experience of the interaction. That a human voice coming from a robot might induce an unpleasant feeling in participants. Results show that the sinusoids got a higher average of 5.11 in comparison to 4.33 for the human sounds. The difference however is not statistically significant ($p > 0.05$).

6. DISCUSSION

From the result could be derived that classic robot sounds, i.e. sinusoids, is perceived as being more natural than human nonverbal expressions when coming from a robot. This indicates a confirmation of our hypothesis, that pop-cultural influences greatly impacts our expectations of what a robot should sound like. The questions about how well the different emotions were conveyed gave inconclusive results. Somewhat of a trend could however be spotted as more people seemed to think that the human sounds were easier to comprehend than the sinusoids, which gives an indication that with a larger test group results may have been more substantial. Evaluating the free text question about sonic expectations, the group that received the human sounds had a tendency to describe the interaction as creepy or scary, which also correlates with the lower score the human sounds received on this question. This suggests that robots generate an uncanny feeling in the observer when producing human-like sounds, which would confirm our hypothesis.

As for the main research question, the natural language of robots could be seen as our perception of how a robot should sound, and therefore something that is constantly changing. Humanoid robots are still widely considered a novelty and one could speculate that as our exposure to this kind of technology increases, our expectations may change, perhaps to a scenario where human sounds is considered as normal in robots. To further explore this would be an interesting topic for future research.

The result can however only be considered true for the rather homogenous test group used in the study, made up by graduate students in media technology. As a group, the test subjects could be considered as having a greater interest in technology than average and also generally having more experience of robot interaction, which may have had an impact on the results. Further tests would have to be conducted to ascertain a more substantial result. As mentioned previously, the size of the test groups could be considered as a source of error as several questions yielded inconclusive results even though the answers pointed towards a difference between the test groups. With a larger test group, a statistically significant difference might have been ascertained.

As previously mentioned, the timing of when Pepper reacted to the story told by the participants was controlled by the test conductors. Therefore, slight variations of when Pepper's reaction was triggered might have occured due to the human factor, which could have possibly affected the participants experience.

This could be the cause of a detail in the results that might deserve mentioning, namely, that the question about how genuine the interaction with Pepper felt got the lowest overall score. Another possible explanation for the relatively low score on this question is that the reactions could be perceived as excessive for some emotions, or possibly the gesture and sound didn't feel congruent. But as suggested by Salem et al [7], incongruent co-verbal gestures did not affect anthropomorphic perceptions negatively, but rather positively. This suggests that questions eight and nine in our survey could have received a higher score than it would have with more congruent gestures.

Another possible source of error is the sound quality of the human emotion expressions. The selected sounds, recordings of a male voice actor with a normal voice pitch, didn't correspond to Pepper's appearance in our opinion. Therefore the recordings were increased in pitch to make a better fit. This altering of the human sounds, as well as the quality of the speakers inside Pepper's head, could have had an impact on participants' experiences of the interaction with Pepper. A more suitable approach might have been to have used a child voice actor instead of an adult, thus possibly creating a more natural experience. This option was however not included in the collection of verified emotion recordings used in the study. As mentioned in the results, some participants expressed the sounds as being mechanical and unnatural, but whether this is a result of the audio in combination with Pepper's appearance and speakers or only the sounds themselves is yet to be determined.

7. CONCLUSION

To conclude, the outcomes of this study suggest a preference of sinusoid sounds regarding the preconceived notions of how a robot should sound. The data also points to expressions of the robot being perceived to be more genuine and pleasant in the sinusoid version of the audio feedback, although this could not be statistically proven. Emotion expressions using a human voice was however slightly easier to interpret by the test group, but this too could not be statistically proven. The results favors our hypothesis, that the influence of pop-culture's portrayal of robots has had a real impact on how we expect robots to sound. Due to the nature of the question, the natural language of robots does not have a definitive answer. As our perception of what a robot is changes, so does our expectations of its sonic representation. In the present state, however, the trend seems to lean towards sinusoids being a more natural candidate than human sounds.

8. REFERENCES

1. Mustafa Suphi Erden. 2013. Emotional Postures for the Humanoid-Robot Nao. *International Journal of Social Robotics* 5, 4: 441–456.
2. Emma Frid, Roberto Bresin, and Simon Alexanderson. Perception of Mechanical Sounds Inherent to Expressive Gestures of a NAO Robot - Implications for Movement Sonification of Humanoids. 10.
3. Markus Häring, Nikolaus Bee, and Elisabeth André. 2011. Creation and Evaluation of emotion expression with body movement, sound and eye color for humanoid robots. *2011 RO-MAN*, 204–209.
4. Adrian B Latupeirissa, Emma Frid, and Roberto Bresin. SONIC CHARACTERISTICS OF ROBOTS IN FILMS. 6.
5. César F. Lima, São Luís Castro, and Sophie K. Scott. 2013. When voices get emotional: A corpus of nonverbal vocalizations for research on emotion processing. *Behavior Research Methods* 45, 4: 1234–1245.
6. James A. Russell. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology* 39, 6: 1161–1178.
7. Maha Salem, Friederike Eyssel, Katharina Rohlfing, Stefan Kopp, and Frank Joublin. 2013. To Err is Human(-like): Effects of Robot Gesture on Perceived Anthropomorphism and Likability. *International Journal of Social Robotics* 5, 3: 313–323.
8. Sichao Song and Seiji Yamada. 2017. Expressing Emotions Through Color, Sound, and Vibration with an Appearance-Constrained Social Robot. *Proceedings of the 2017 ACM/IEEE International Conference on Human-Robot Interaction*, ACM, 2–11.
9. Pepper the humanoid robot | SoftBank Robotics EMEA. Retrieved October 16, 2019 from https://www.softbankrobotics.com/emea/en/pepper.
10. Choregraphe Suite — Aldebaran 2.4.3.28-r2 documentation. Retrieved October 17, 2019 from http://doc.aldebaran.com/2-4/software/choregraphe/index.html.
11. Pure Data — Pd Community Site. Retrieved October 14, 2019 from https://puredata.info/.

Appendix

Before you start reading the stories, say hello to Pepper

Story 1:

When the last gorilla at the zoo died, the owner asked one of her workers to wear a gorilla suit and pretend to be a gorilla. To get people's attention, the worker climbed over the lion's cage but lost his grip and fell. He started screaming "HELP!". Then the lion punched him and whispered: "Shut up or you're gonna get us both fired!!"

Story 2:

When I was a child, I had a cat named Scarlet. She was my best friend, and she always greeted me at the door when I came home. One day, when I opened the door, Scarlet wasn't there. I got worried and went looking for her at her favourite spot, beneath the oak tree in the backyard. There she was, laying completely still. The next day we had her funeral beneath that same oak tree.

Story 3:

Yesterday, I was eating lunch with a friend when my phone rang. I had been waiting for a call from the hospital about my sister who is sick. When I answered, the nurse sounded distressed, so I got worried. Then she told me that my sister was feeling much better and she just wondered when I could come and pick her up!

Story 4:

Once upon a time there was a mechanic looking for scraps at the scrapyard. Suddenly a huge monster made out of old car parts stood before him. The mechanic was really scared and begged: "Please don't hurt me!
The monster said: "Don't worry, **I ONLY EAT ROBOTS**!"