Building Believable Robots

An exploration of how to make simple robots look, move, and feel right

by

Paul Bucci

B.A. Visual Arts, The University of British Columbia, 2012

B.Sc. Computer Science, The University of British Columbia, 2015

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

Master of Science

in

THE FACULTY OF GRADUATE AND POSTDOCTORAL STUDIES

(Computer Science)

The University of British Columbia

(Vancouver)

August 2017

© Paul Bucci, 2017

Abstract

Humans have an amazing ability to see a ‘spark of life’ in almost anything that moves. There is a natural urge to imbue objects with agency. It’s easy to imagine a child pretending that a toy is alive. Adults do it, too, even when presented with evidence to the contrary. Leveraging this instinct is key to building believable robots, i.e. robots that act, look, and feel like they are social agents with personalities, motives, and emotions. Although it is relatively easy to initiate a feeling of agency, it is difficult to control, consistently produce, and maintain an emotional connection with a robot. Designing a believable interaction requires balancing form, function and context: you have to get the story right.

In this thesis, we discuss (1) strategies for designing the bodies and behaviours of simple robot pets; (2) how these robots can communicate emotion; (3) and how people develop narratives that imbue the robots with agency. For (1), we developed a series of four robot design systems to create and rapidly iterate on robot form factors, as well as a tools for *improvising* and *refining* expressive robot behaviours. For (2), we ran three studies wherein participants rated robot behaviours in terms of arousal and valence under different display conditions. For (3), we ran a study wherein expert performers improvised emotional ‘stories’ with the robots; also one of the studies in (2) included soliciting narratives for the robot and its behaviours.

Lay Summary

Humans have an amazing ability to see a ‘spark of life’ in almost anything that moves. It’s easy to imagine a child pretending that a toy is alive. Adults do it, too. Leveraging this instinct is key to building believable robots, i.e., robots that act, look, and feel like they are social beings with personalities, motives, and emotions. Although it is relatively easy to make a robot seem alive, it is difficult to control, consistently produce, and maintain an emotional connection with a robot. Designing a believable interaction requires balancing form, function and context: you have to get the story right. For this thesis, we explored how robots can communicate emotions through simple body movements. We found that people will perceive a wide range of emotions from even simple robots, create complex stories about the robot’s inner life, and adapt their behaviour to match the robot.

Preface

This thesis is organized around two major published papers, with some current and unpublished work. For all work, I collaborated closely with members of the SPIN Lab, especially Laura Cang, Oliver Schneider, David Marino and my supervisor, Karon MacLean, and two visiting researchers, Jussi Rantala and Merel Jung. In conjunction with Laura and Oliver, I supervised a number of undergraduate researchers, all of whom contributed time and effort to the projects, including Sazi Valair, Lucia Tseng, Sophia Chen and Lotus Zhang. I am extremely grateful to each of them, however, I would attribute most of the intellectual contribution to myself, Laura, Oliver, and David. In *CuddleBits* [16], I was the principle designer, architect, and builder of the CuddleBit form factors and design systems. Although other people helped with evaluation and assembly, this is my work. Further CuddleBit designs included a twisted string actuator introduced and developed by Soheil Kianzad.

The chapter on *behaviour generation* (4) outlines a number of behaviour generation ideas and systems. *MacaronBit* is an extension of work by Oliver that was initially developed by myself, Oliver, Jussi, Merel, and Laura, and carried into maturity by myself and David. *Voodle* [51] is a system initially conceived of by Oliver and prototyped by Oliver and David. I architected and implemented Voodle's final versions that were used in our studies, with significant help from David in writing the final code. The *Twiddler* and *Hand-Sketching* prototypes were developed on my own. The work in *Complexity* stems from ideas developed by Laura and myself in *CuddleBits*, but was carried out to maturity by myself, with significant help from Lotus in taking care of the very important minutia.

The chapter on *behaviour evaluation* (5) outlines a number of studies in which we evaluated the CuddleBits and Voodle, through both displayed and interactive behaviours. In *CuddleBits*, the initial study design was developed collaboratively between Laura, Oliver, Jussi, Merel, and myself, then brought to maturity by Laura and myself. In *Voodle*, I largely architected the study design, with help from David. *Voodle* analysis was performed collaboratively between myself, David, and Oliver.

As first author on *CuddleBits*, I principally drafted, organized, edited and wrote the paper. Laura and Karon contributed significantly to editing, with help from Jussi, Merel and Sazi.

On *Voodle*, I contributed heavily to initial drafting, framing, analysis and writing. David, Oliver, and Karon carried the paper through the first draft, then I edited and wrote heavily for the second draft. The final draft was largely an effort by Karon and David.

The chapter on *current and future work* (6) includes ideas that are currently in development by myself, Laura, and Mario Cimet in an equal intellectual partnership. It also references unpublished work developed by Laura, Jussi, Karon, and myself, with significant inspiration from Jessica Tracy.

The chapter on *related work* (2) is largely taken and amended from *Voodle* and *CuddleBits*. Sections on *believability* and *complexity* are my own.

Research was conducted under the following ethics certificates: H15-02611; H13-01620; H09-02860.

Table of Contents

List of Tables

List of Figures

Acknowledgments

No good work is done alone. I certainly couldn't have done anything interesting without support from my colleagues, friends, and loved ones. It is rare that we allow ourselves the opportunities to be effusive, so I will do my best to make the most of it.

To each of my colleagues at the SPIN Lab and MUX group, I thank you for your critique, enthusiasm, and support. I should mention a few by name:

Laura, for over two and a half years of productive collaboration and friendship. It's a rare thing to find someone who is able and willing to take one's ideas seriously. You've done that and much more. I am very grateful to have worked with someone so enthusiastic, intelligent, and thoughtful.

Oliver, for consistent mentorship and enthusiastic help whenever I needed it. I learned a lot from you about research, and I'd like to say I taught you a thing or two as well. I've enjoyed getting to know you, and I hope we can work together again.

David, for listening to me talk for hours, encouraging me, working with me, writing with me, tolerating me, and believing in my silly ideas. It's been a privilege and a joy to work with you and be your friend.

Jussi and Merel, for a very creative half a year and excellent intercontinental collaboration for the next couple of years.

Karon, for believing in me many times over, pushing me forwards, looking carefully and deeply at my work, and saying 'yes' more than you needed to. It's been a privilege to work with you and learn from you.

This work was also supported by contributions from many people in other departments. Jon Nakane (EngPhys), Nick Scott, Graham Entwistle and Blair Satterfield (SALA) deserve special mention.

I'm personally grateful to Tamara Munzner, Christine D'Onofrio, and Jessica Dawson, all of whom have given time and thought to helping me advance in my academic career. And to Eric Vatikiotis-Bateson, who believed in me and helped shape my perspective on science. RIP.

And, of course, I'm deeply grateful my friends and family who have supported me along the way.

Ashley, who has been willing to listen to hours of my fool ideas for years now.

All of the Oldbyssey and Abby crews who have politely engaged me in conversation about this.

Micki, for love, and for putting me up while I finished this thing.

My mother, Dorothy, who has spent many car rides, late nights, and early mornings listening to me talk, ever since I was able to talk.

Dedication

To Micki, who has shown me more kindness, support and love than I have yet known.

Chapter 1

Introduction

The most sophisticated people I know—inside they are all children.
— Jim Henson

This work is pretty silly. I want, you, the reader, to know that up front, and to know that I know that this is silly. But I have taken this silly stuff pretty seriously, and have found that keeping things playful has helped me and my colleagues tackle some big and interesting problems in affective computing and social robotics. So, dear reader, prepare yourself for a monograph that looks seriously at crafts, toys, puppetry, and funny noises, and hopefully you'll agree that being a little silly can produce some interesting results.

A note on pronouns: "I" will refer myself, i.e., the author of this document, Paul Bucci. "We" will refer to myself and my collaborators in the Sensory Perception and Interaction Research Group, a.k.a. SPIN Lab.

1.1 What Makes a Social Robot Believable?

Deciding what constitutes a robot can be difficult. If you go broad, any interactive system can seem like a robot. Do you include wearable electronics? How about an interactive light installation? Does the robot need to have a body? Can you call a character in a cartoon a robot? Does the robot need have senses, i.e., vision, hearing, touch?

Here, I'll be discussing *social robots*, which are robots that need to act in a

Figure 1.1: Examples of social robots. The Nao (left) is a humanoid robot, Google's self-driving car (centre) acts in human social spaces, and Universal Robotics' UR-10 is a collaborative pick-and-place robot.

human social space. You can imagine robots that greet you at the door like the Nao[1], or that you collaborate with like many pick-and-place robots[2], or that drive you around like Google's self-driving car[3] (see Figure 1.1).

For this work, we'll assume that, yes, a robot does need to have a body. That leaves on-screen cartoon characters out of the robot category. We'll further assume that a robot needs to be perceived as an *autonomous agent*, or an independent moving thing with a mind of its own. Interactive light installations, automatic doors, and wearable electronics are all out, too. Finally, we'll assume a robot needs to be able to sense and react to the world, so, stuffed animals, cuckoo-clocks and wind-up toys are out.

I'll confess upfront to cheating a little. Although the robots described herein can and have been given both sensing and reactive capabilities, some of the studies we ran had robots that could not sense and could only act. But I would still be confident in calling them believable social robots—why?

As an exercise for the reader, try ordering the examples below in terms of robot-ness, then social-ness[4]. You can decide whether you agree with my ordering. The point is not to determine some universal order, but rather to try to tease apart what makes something *feel like a social agent*.

[1] https://www.ald.softbankrobotics.com/en/cool-robots/nao

[2] Example pictured from https://www.universal-robots.com/

[3] https://waymo.com

[4] I am intentionally using the *-ness* suffix over robotic and sociality to avoid connotations of repetative, automatic movements and ability to socialize, respectively.

Try it.

- on-screen cartoon characters
- interactive light installations
- automatic doors
- wearable electronics
- stuffed animals
- cuckoo-clocks
- wind-up toys

Done? Both social-ness and robot-ness? My turn. I'll just do three. I would order *stuffed animals* $<$ *automatic doors* $<$ *wind-up toys* in terms of robot-ness. But if I were to order in terms of social-ness, I would order *automatic doors* $<$ *wind-up toys* $<$ *stuffed animals*.

Automatic doors, although they live in a human social space, don't interact emotionally with the humans around them. Although humans very often get emotional at automatic doors—especially if they don't open when expected—the doors aren't seen as malicious, just poorly-designed. In other words, we don't develop a *theory of mind* by attributing desires, emotions, or intents to the doors; we think of them just as machines.

Wind-up toys seem to have a little more agency. They move on their own, seem to have a purpose (even if a little single-minded), and they can evoke emotions. Automatic doors are rarely funny, but wind-up toys are often funny, surprising, and even cute. As toys, they are designed to be played with, which puts you in a frame of mind where you're willing to pretend that the interaction is real in some way, despite your knowing that the the toy is a toy. This idea of *suspending your disbelief* comes from theatre: it's your ability to become immersed in the emotions and narrative on the stage despite the fact that (a) you know it's a play; (b) the actors aren't acting like real people; (c) the set doesn't look like the real world. We can become fully immersed in a pretend-world even when there is a single actor by themselves on a featureless black stage, as long as we are willing to suspend our disbelief about how unreal the play-acted scene is. In fact, sometimes adding more features to a set becomes a distraction. Adding some *thing* to a scene filters the possibilities for imagination.

This *paradox of presence*[5] is what ranks a stuffed animal above a wind-up toy in terms of social-ness. Stuffed animals don't move on their own, but they are also not constrained to a single kind of motion through a fixed mechanism. The deterministic aspect of the wind-up toy takes away from all of the emotions you can imagine it having. Stuffed animals, on the other hand, are free to be imbued with whatever story you want to give it. If you're willing to play pretend and to suspend your disbelief, a stuffed animal can feel very real indeed, even if you know rationally it's not. Watch a child play with their stuffed toys one day—it's obvious that their pretend scenes with their toys are real enough to them to impact their emotions.

As a final thought experiment, imagine what it would take to make an automatic door seem like it had agency. It would have to occasionally act against the purpose the designer intended for it[6]. An automatic door that opens courteously is just a door that works well. But an automatic door that played games with you—tried to close on you halfway through, started to open and then slammed shut, teased you by remaining open when you were far away but closed as soon as you tried to enter it—would seem malicious. You would imbue it with (an unfortunate) theory of mind because you would be unable to predict its actions well.

The above examples illustrate three key elements for treating an object like a social robot with agency:

1. The object's ability to behave autonomously (can it move on its own?).

2. How deterministic the behaviour is (can you predict what it will do?).

3. Your willingness to suspend your disbelief, which is dependent on the social

[5] A soft paradox: adding things to a set seems like it shouldn't take away from an experience, it should add to the experience. But in this case, adding some thing to the set takes away from what you can imagine. This is the basis of the *more is less* aphorism.

[6] If you can think of any examples where the door both works well and seems to have agency without dipping into teasing or some other faux-malicious act, please let me know.

situation of the object as well as its abilities and behaviours (how much do you want to play along?).

A believable social robot, therefore, is something you have to act with *as if it were a social agent*, regardless of how much you have to suspend your disbelief. This puts the burden on you, the interactor[7], as it is *your frame of mind* that produces the robot as believably social more than the robot itself. The magic moment that elevates a robot from machine to agent happens *in concert with the robot*, but ultimately it is up to you to decide the robot has a mind of its own. As the rest of this thesis will argue, that sense of agency seems surprisingly easy to produce, but very difficult to control.

1.2 Affective Touch and Narrative Context

Humans communicate through a number of channels: speech, touch, posture, writing, drawings, to name a few. Intuitively, we could organize the way meaning is conveyed into *explicit* and *implict* semantic paradigms. For explicit semantics, meaning is conveyed through signifiers such as words or pictures, and there is some formal logical content to the statement. For example, if I told you “your dog ran away,” you would be able to derive some verifiable truth value regarding what happened to your dog. However, implicit semantics convey meaning through context and inference. For example, if I told you “your dog ran away” with a dejected tone of voice, you might infer that the situation was dire, that I didn’t know where your dog might be, and you just might not be getting your dog back. But if I told you the same thing with a neutral or hopeful tone of voice, you might infer that I might be following up with “...and we found it,” or “...but I have a pretty good idea of where it is,” and there might be some hope for recovery.

Touch interaction affords both implicit and explicit communication. Consider teaching someone how to dance, specifically trying to correct their posture. Imagine trying to describe in words how each body part should be oriented or which muscles should be activated. Anybody who has worked as a dance instructor or a

[7] Here I use *interactor* rather than *user* to emphasize the bidirectionality of the interaction and the activity of the interaction, i.e., inter-actor. This also attempts to understand the person commonly referred to as a *user* as a more complex and interesting human being.

personal trainer will tell you how difficult that task is. It is nearly impossible to give them a series of instructions that describe how they should amend their exact posture from moment to moment. However, by placing your hands on them and pushing their body into place, you can fully express how you want their posture to be with very little effort and no words. This would be explicit instruction via touch.

Contrast to how a lead communicates with a follow during a partner dance. Usually, there is a slight touch that might suggest the move that the lead expects the follow to make next, but the follow must use their knowledge of the dance, the part of the song that they're in, and their knowledge of the lead's abilities to infer the lead's intended next move. This constructing of knowledge through inference motivated by a touch would be an implicit instruction via touch.

There is also a deep emotional dynamic communicable through touch. You can covey affect (i.e., emotion) powerfully through touch: imagine the difference between saying to someone "I love you," and giving them a full-bodied hug. A certain depth diversity, power and depth of meaning can be expressed through touch. Imagine the scenario where your partner is about to do something rash. You could explicitly tell them "don't do that," with a tense tone of voice, which tells that you want them to stop. Or you could touch them softly but firmly on the arm, which conveys the same thing, but also closeness, care, and trust. Which do you think would work better?

The field of affective touch studies this interesting space of implicit communication. Although there can be explicit meaning in a touch gesture, the context of the interactors greatly impacts the content of the meaning. In linguistics, context-dependent meaning creation is studied in the subfield of *pragmatics*. Although I won't be going deep into pragmatics in this work, it is important to keep in mind this interplay of context and meaning. Specifically, I consider the relationship between the robot and the interactor, how the interactor sees the robot, and the story that the interactor builds about the robot. Together, I refer to this as the *narrative context*, which frames the interaction with the robot as *storytelling*. This borrows from a long tradition in the social sciences of considering identity creation and interaction as a narrative act. The term is meant to invoke the literary connotations of a narrative, i.e., characters, themes, story arc, point of view. As will be argued later, it is important that an interactor develops a narrative context for the robot by deciding which metaphors represent the robot (Is this robot a cat? Dog? A mix?), the interactor's role (Am I an owner? A friend? Neither?), a history of their relationship (Did we always get along?), and a theory of mind for the robot (What does the robot want or believe?).

1.3 CuddleBots and Bits

A number of years ago, a lap-sized furry robot called the Haptic Creature was developed in the SPIN Lab to study affective touch. Resembling a sort of mix between a cat and a guinea pig, the Haptic Creature was imbued with the ability to *display* emotionally-expressive movements through inflatable ears, an expanding ribcage to simulate breathing, and a vibration motor to simulate purring, and *detect* touches through a matrix of force-sensitive resistors (FSRs). After SPIN researchers showed the Haptic Creature to have a measurable physiological and emotional impact on study participants by calming them down through breathing behaviours, the lab developed a new version of the robot called the CuddleBot. With this robot came the challenge of impacting interactor's emotions in other directions, such as making them more excited, stressed, or depressed. The dream is to make interactions with the robot range a full spectrum of emotion, and, hopefully, make it possible to create emotional scenes with the interactors that could enable entertainment, education, or therapy.

However, as the robot was developed and used, it became clear that a traditional engineering-focused robot design process would not work for emotional expression. As such, we developed small, 1-degree-of-freedom (1-DOF) robots that explored a single expressive channel at a time called the Cuddle*Bits* (as in the diminutive of multi-DOF Cuddle*Bots*). Critically, these robots enabled rapid iteration on their form and abilities, a process we refer to as *physical sketching* [57]. Further, we developed tools for designing the robot's behaviours, including an improvisation tool that enabled *behaviour sketching* and an editing tool that enabled *behaviour refining*. By taking this improvisational sketching approach, we were able to quickly iterate on the robot's bodies and behaviours, developing both aspects simultaneously for a cohesive exploration of the design space.

1.4 Approach

Targeting believable emotional interactions as the long-term goal of the CuddleBot project, in this thesis we focus on the development of the CuddleBits, their coupled body and behaviour design, and their evaluation in terms of emotional expressivity. Using qualitative and quantitative approaches, we submit that the CuddleBits (a) are capable of consistently expressing a wide range of emotions; (b) allow interactors to develop a sense of alignment and agency under certain conditions; and (c) illustrate a viable design approach for developing robots capable of emotion expression.

For (a), we ran three studies where participants were asked to rate displayed robot behaviours in terms of pleasantness (valence) and activation (arousal). For (b), we ran a three-part co-design study over six weeks with expert performers. For (c) I designed a series of robots, wherein I developed and used a sketch-based rapid prototyping approach for emotionally-expressive robots.

1.5 Contributions and Limitations

This thesis is comprised of three major works, referred to as *CuddleBits* (Simple Robots for Studying Affective Touch), *Voodle* (**V**ocal **Doodl**ing for Improvising Robot Behaviours), and *Complexity* (Characterizing Pleasant/Unpleasant Robot Behaviours). The contributions of each are outlined here; since all share similar domains, limitations are summarized for all works at the end.

1.5.1 CuddleBits: Simple Robots for Studying Affective Touch

In *CuddleBits*, we devised a rapid prototyping approach that allowed us to design and quickly iterate on a small family of 1-DOF robot bodies (Figure 3.1). The robots were designed *with sketching/modification as a primary requirement*, thus allowing for iteration on the order of under a minute to less than a day. Simultaneously, we explored *design tools for robot behaviours*, extending a keyframe-based editor for vibrotactile sensations (Macaron [79]) and a new tool for improvisational haptic sketching with vocal input (Voodle).

We developed, tested and analyzed a set of emotion behaviours. Study participants rated the emotional perception of our behaviour set for agreement with validated emotion words. We analyzed the behaviour set for design parameters and characteristic signal features (mean, median, max, min, variance, total variance, and area under the curve) to determine which aspects of robot motion correlate with arousal and valence. We further examined robot form factor and participant perception of emotion under two sets of two conditions: surface {soft and furry vs. hard and exposed}, and viewing condition {live vs. video} in a mixed experiment design.

We contribute:

- *DIY designs* for a 1-DOF robot with a validated range of emotional expression;
- Identification of *relationships* between robot behaviour control parameters, and robot-expressed emotion (as consistently rated by participants).
- *Demonstration of the capability and potential* of a sketch-and-refine design approach whose novelty lies in facilitating joint consideration of form and behaviour for 1-DOF affective robots, and can be extended to more complex displays.

1.5.2 Voodle: Vocal Doodling for Improvising Robot Behaviours

In *Voodle*, our goal was to support creation of believable robot behaviour through vocalizations. We developed voodling in stages: (i) a pilot exploration of vocal interaction with a linguistic analysis, to test the concept and gather requirements; (ii) a comparative study with 10 naïve users to situate Voodle in relation to traditional animation tools (also reported in *CuddleBits*); and (iii) a co-design study with three expert performers in a 6-week intensive relationship while the Voodle system was iteratively revised. We describe Voodle's design and implementation, then detail and reflect upon our two studies.

We contribute:

- A *working Voodle system* that is customizable in real-time, and extensible for further development and applications.

- *Key factors underlying effective voodling* in affective interaction, relating to level of user control, form, and achievable alignment and believability.

1.5.3 Complexity: Characterizing Pleasant/Unpleasant Robot Behaviours

With *Complexity*, we explored the relationship between the complexity of a behaviour-producing waveform and the perceived valence. Building on our previous observations of how interactors create 'stories' for the robot, we also inspect how narrative framing can invert how the same behaviour is seen by different people.

We contribute:

- An analysis of *complexity measures and valence* that show how valence and particular complexity measures correlate.
- An analysis of *narrative framing* in relation to valence perception, with an emphasis on understanding how the robot is seen to have an 'inner life' and agency.

1.5.4 Limitations

The major strength of this work is also one of the biggest limitations: these robots are 1-DOF by choice, which limits how much they are able to express certain emotions. All claims made herein, therefore, should be understood to be relative to these very simple robots. Any extrapolations beyond 1-DOF are purely speculative.

This includes the exploration of design systems. It is not yet known how well our design systems will extend into two or more DOFs. As such, the contributions in design systems should be seen as enabling faster creative expression *only within the prescribed limitations of a particular design system*. As such, this is more of a call to action: we espouse the design approach of treating social robot design as primarily artistic enterprise, and apply known practices and lessons of art and design production to robot design.

Our claims made in reference to the robot's emotion expressivity are similarly made only relative to the emotion models we used, which are simplistic and have known limitations (discussed in *Related Work*. Critically, the models do not account for narrative point of view, the possible non-linearity of emotional space, and how measurement tools are culturally situated. As such, it is important to think of the experiments and quantitative results discussed herein as rhetorical devices used to provide insight into social phenomena.

1.6 Thesis Organization

This thesis is organized around two major papers and one unpublished work. The content of the two papers has been reorganized to reinforce the central theme of building believable robots. *Related Work* outlines concepts relevant to all works discussed here, *Sketching* presents the physical design of the robots, *Generating* presents our work on producing robot behaviours, and *Evaluating* presents our work on evaluating robot behaviours. *Current and Future work* presents work that is on-going in the SPIN Lab to create a model of interactor's emotions, critiques current approaches, and outlines future directions for research. This work has been developed collaboratively, as noted in the preface.

Chapter 2

Related Work

To enable emotional interaction, machines must be able to both *sense* and saliently *display* emotions [31]. Here, we focus on the display of emotion through breathing-like behaviours. We root our exploration of affective robot motion in the literature, borrowing insight from companion robots' emotional breathing, and considering available emotion models.

We take inspiration from other places where physical and behavioural design is coupled, discuss the role of improvisation in the creative process, outline the inspiration for and implementation of our custom animation methods and tools that underlie our motion rendering process, and end with a discussion on what makes robot interactions believable.

2.1 Companion Robots, from Complex to Simple

Companion robots that once existed only in science fiction are quickly becoming part of our present reality. Paro, a cute actuated harp seal with soft fur, has been used as a therapy robot in elder care homes to help manage dementia and encourage socialization [88, 89]; its study provides evidence that even simple robot behaviour can produce therapeutic benefits.

The Haptic Creature is a furry mammal-like robot with a multi-DOF haptic emotion display [95], shown to display a variety of emotions with multi-DOF [96]. The physiological effect of one level of sinusoidal robot breathing was studied in a controlled trial, where users held the robot in "breathing" and "non-breathing" conditions. In the first, they experienced a statistically significant decrease in their own heart- and breathing-rates as well as self-reported relaxation [81].

Here we examine the larger *expressive* space of breathing-like behaviours alone.

2.2 Rendering Emotion Through Breathing

Breathing is a natural expression of emotion [11, 12, 64] that can enhance the recognition of certain affective states, e.g. as displayed on a 3D avatar of a human upper body (including the face) [28]. It is yet unclear how to operationalize design and/or display parameters to distinguish between emotional states. Boiten et al. identify two relevant categories of respiratory parameters: (1) volume and timing parameters (i.e., respiration rate and respiration volume, frequency and amplitude respectively), (2) measures regarding the morphology of the breathing curve (i.e., changes of the breathing curve over time; irregularity) [12]. Our keyframe editor, MacaronBit, attempts to capture both sets of parameters; our improvisation tool, Voodle, manipulates these parameters implicitly.

Keyframes come from animation, where the *key frame* is the animation frame where some motion changes. The animator will set the value of some variable at two points in time—say the height of a bouncing ball to 1 meter at 0:10 seconds and 0 meters at 0:12 seconds—and the computer will fill in the animation frames between the two keyframes.

For volume and timing parameters, imagine a person breathing regularly. If they breathe deeper and faster, you would assume they are more excited. For morphology of the breathing curve, imagine you draw a regular breath as a graph: this would look something regular and periodic like a sine wave. Now, imagine drawing someone crying. The graph would look a lot more erratic, with a lot of jumps and stops randomly distributed on the curve.

In previous work on similar 1-DOF robots, people petting the small furry bodies reported that their simple periodic motions seemed analogous to breathing, suggesting aliveness [15, 18]. Their behaviours have been likened to other repetitive biological behaviours such as wing motion or heartbeats.

2.3 Emotion Models

Modelling human emotions presents many difficulties, and emotion model usage is fraught with misunderstanding and misuse [94]. This is because there are a lot of difficulties in measuring emotions. Emotions are felt privately by the person who is experiencing them, and there exists no truly objective measure without being able to completely model how someone is thinking. Physiological measures such as skin conductance or heart rate can give you an objective measure of the effects of some set of emotions, but you still have the problem of matching the subjective emotions to the objective physical effects. As such, emotion researchers make up methods to measure emotion by using scales based on emotion words, such as "happy" or "sad," then ask study participants to rate their feelings in terms of those words ("how happy do you feel from one to five?"). But there is a mismatch between the theories of emotions are actually produced in the brain and these kinds of word-based models (called "categorical" models). Some theories of emotion attempt to organize all possible emotions onto abstract dimensions such as "pleasantness" or "activation" (called "dimensional" models) and then try to place emotion words along those dimensions.

Categorical emotion models attempt to arrange the full spectrum of human emotion into discrete categories, often represented by one or more emotion words (such as "excited" or "stressed"). Structurally, categorical emotion models do not account for conflicting or multiple emotion states (you can feel both "happy" and "sad" at the same time), and are not granular enough to account for the subtlety of felt experience (there are many kinds of emotions we might group under "happy"). These models are based on being able to consistently describe an emotion with words, which introduces many difficulties with individual and cultural differences.

Dimensional models attempt to present emotion experience as continuous along supposedly orthogonal dimensions such as "pleasantness," "activation," or "dominance." These models have the same shortfall of being unable to represent multiple simultaneous emotion states[1], but are by definition highly granular. However, two points can be very close dimensionally, but far semantically: in a 2-dimensional

[1]Unless a measurement is treated as probabilistic—for example, you could imagine using a dimensional model as a probability density function, where you distribute weight to every point on the emotion space. However, this is not how they are typically used.

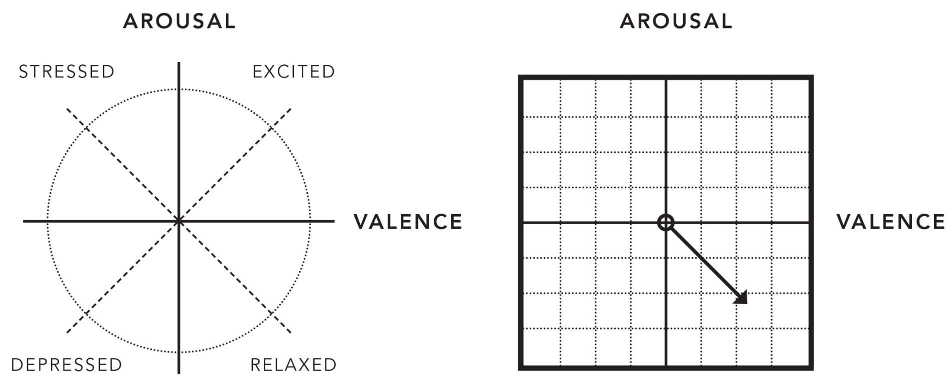


Figure 2.1: Two dimensional models: left, Russell's circumplex with associated PANAS words; right, the discretized affect grid.

model that uses "pleasantness" and "activation", both *angry* and *nervous* are spatially close (both are highly activated and unpleasant), but are conceptually very different[2]. To be useful to researchers, dimensional models are often discretized by either directly portioning the space (such as with an affect grid, see Figure 2.1) or using categorical analogues (such as with the Positive and Negative Affect Schedule, or PANAS, where words are mapped to grid points). This reintroduces the problems with categorical scales, but with an attempt to sidestep the dependence on language. The argument is, roughly, if the core dimensional model can be shown to be generalizable, a discretized version of that model might represent a more valid construct.

Since each person's subjective frame is variable enough that without heavy calibration, qualification, or verification, an objective measure can over-represent the referent subjective state as true. That is not to say that any attempt at quantifying subjective states is doomed to failure, but that the use of a quantified scale needs to be understood as fundamentally rhetorical and relationally constructed.

We use a conventional two-dimensional emotion model of valence (pleasantness) and arousal (activation) [71]. This model is often discretized by dividing the 2D plane into a grid [72]; the Positive and Negative Affect Schedule (PANAS) [92] is a set of words with validated mappings to divisions of the resulting affect grid [30].

[2]3-dimensional models that add "dominance" claim that this apparent closeness is an artifact of projecting into 2-dimensional space, but there are counterexamples even in a 3-dimensional space.

We use a subset of PANAS words to refer to the grid's extreme corners: *Excited* (positive valence, high arousal), *Relaxed* (positive valence, low arousal), *Depressed* (negative valence, low arousal), and *Stressed* (negative valence, high arousal), as well as others (a total of 20 words) that fall into each quadrant. This model choice is motivated by dimensionality, with future behaviour generation on continuums of valence and arousal in mind. However, we also take care to calibrate word usage, verify the representation of emotion states by checking in with participants, and account for multiple possibly-conflicting emotion states through using multiple and complex ratings.

2.4 Complexity and Physiological Behaviours

In biological, computational, and artistic signal processing, there is a need to differentiate *complexity* from variability[3] [44, 50]. To develop an intuition for this, consider how randomness is perceived. Imagine a T.V. screen displaying pure static—i.e., white noise. If you were to split that screen up into a 10 by 10 grid of boxes, you could easily switch two boxes and perceive no noticeable difference in the picture. Static looks the same everywhere. White noise is uniformly random, which means that at every point on the screen, it is equally likely that you will see any of the possible values a point could have.

By contrast, imagine a black and white splatter painting such as the ones that Jackson Pollock painted. Again, split the splatter painting into a 10 by 10 grid of boxes. Now, if you switch two boxes, it will be really obvious that you've made a switch, even standing very far away and cutting/pasting the switched boxes perfectly. Certain paint splotches will look discontinuous, and you could very easily see the sharp lines along which you made the cut.

Both are random, in some sense of the word. But the static is more random in the mathematical sense, and the splotchy painting is more random in the perceptual sense. In a way, the uniformly distributed white noise has no structure to it. You can't see any patterns except total randomness. Therefore, we would call the splotchy painting more perceptually complex than the screen full of static [1].

If you still need convincing, think about mirroring the image of the static vs.

[3] No relation to computational complexity.

mirroring the image of the splotchy painting. Imagine you are shown a picture of the original static screen and splotchy painting, then told to close your eyes while the pictures were randomly swapped out with either the same image or a mirror image. If then asked whether the static screen was mirrored, could you tell? Probably not. But you could easily tell whether the splotchy painting was mirrored.

The "switch two boxes from a grid" approach is called self-similarity. The "guess if the new image is a mirror image" approach is called (time-)irreversibility, i.e., to what extent you can reverse a signal and have it still look the same. Both are measures of complexity: we would say that a more complex signal is less self-similar and less time-reversible [26]. Often, signals that *feel* random are not uniformly random (static, white noise, etc.), they are more complex (paint splotches, Perlin noise, Brownian motion, etc.).

Now, consider some different 1-D time series: a sine wave, white noise, and 1-D Brownian motion. By many measures of variability, the white noise is the most variable, i.e., it has a uniform spectral power distribution, highest Shannon entropy, and infinite variance[4]. However, white noise is also highly self-similar, time-reversible and uniformly random. Intuitively, you should be able to take a slice of white noise and replace it with another slice without perceiving a difference in the signal. In contrast, mixing slices of 1-D Brownian motion would create very different-seeming signals.

A measure that accounts for self-similarity at different time scales is *multi-scale entropy* (MSE, explained here intuitively, below algorithmically) [25]. It is used in the biological sciences as a complexity measure, e.g. to determine how irregular a heart beat is. MSE is defined by a process of taking the *sample entropy* [68] of a signal at different time scales. Sample entropy is a measure of how self-similar a signal is at a single time scale. It is calculated by splitting the signal up into vectors of length k, counting the number of vector pairs that are within some distance r, repeating the count for vectors of length $k+1$, then taking the negative log ratio of the two counts.

Here, we develop a working hypothesis that valence and complexity are negatively correlated for breathing behaviours, i.e., the more complex a breathing be-

[4]Over an infinite signal.

haviour is, the lower valence it is. The intuition comes from our own observations, as well as Bloch [11] and De Melo [28], where breathing behaviours described as irregular were also deemed to be lower valence.

2.4.1 Sample Entropy

Sample Entropy is a measure of self-similarity for some signal, S. The algorithm is as follows:

1. Break S into a list of k-sized vectors, X.

2. Given any distance function, d, and some minimum distance, r, count the number of pairs of vectors such that $d(x_i, x_j) < r$. Call this count N.

3. Now split S into a list of $k+1$-sized vectors, Y, and repeat (2) for Y. Call this count M.

4. Calculate $-log(N/M)$.

2.4.2 Multi-Scale Entropy

Multi-scale Entropy improves on Sample Entropy by looking at different coarse-grained time scales. The algorithm is as follows:

1. Split a signal S into k parts.

2. Average each part to produce a new vector, A, where A will be of length k.

3. Determine the Sample Entropy of the new signal, A.

4. Repeat for all lengths k from 1 to n, where n is the length of the orignal signal, S.

2.5 Coupled Physical and Behaviour Design

The way that a robot is shaped changes the way that it can move. Emotional behaviours are intrinsically bound to the bodies on which they are expressed. Being able to flexibly design and realize robot bodies is imperative to quickly and creatively explore the possible design space for a new robot. As such, affective robot design benefits strongly from the flexible, rapid-turnaround physical prototyping methods being popularized by DIY and maker culture.

Our sketch-and-refine process draws on low-fidelity haptic sketching, traditional sculpture and paper craft, rapid prototyping, and animation methods. Our innovation is a process that integrates existing sketching methods by preparing them in form amenable to quick combination and refinement.

2.5.1 Sketching Physical Designs

As a design approach, *sketching* means quickly generating many versions of an imagined final product. For some engineering problems, detailed specifications can be generated without sketching. In contrast, establishing requirements through fast iteration (design-by-prototyping) [48, 52] is typical for those who aim to generate emotional reactions — e.g., industrial designers, set designers and sculptors, who often develop a larger design by making and getting reactions to multiple simultaneous *maquettes* (small physical models).

Our physical design methods are inspired by Moussette's haptic sketching [55, 56], with complex haptic expressions mocked up with low-cost physical media, and by Hoffman and Ju [40], where robots were designed through drawn sketches, prototyping actuated skeletons, and 3D modeling.

We take "robot sketching" further. Concerned with hapticness as well as visual emotiveness, we prioritize rapid access to user responses to tangible, physical media. Therefore, skipping 3D modelling, we directly implement our robots in low-cost sketch media (wood, plastic) so they are both easy to iterate on and immediately study-ready. We thus design both *for* and *through* sketching: by making on-the-fly modification a primary design requirement, haptic (touchable) robot prototypes can be iterated on within hours or even minutes (Figure 3.1).

In keeping with Maker ethos, we also target open hardware design, inspired by projects like WoodenHaptics [32].

2.5.2 Creating Expressive Movement

Designing expressive behaviours can be challenging, requiring animation, behavior and robot expertise as well as diverse tools [36]. Conventional robot movement is produced by an algorithm that acts on a model to define an exact path towards a goal, optimizing efficiency or safety [21].

Affective robot control differs from typical robot motion planning, with a goal of communicative rather than functional movement. Affective robot behaviour design consequently draws heavily on puppeting (3D) and animation [67]. Our minimal sketch-and-refine approach stands in contrast to those for higher-DOF affective robots such as Probo [75], where focus on facial expression for emotional display necessitates extensive 3D simulation software to coordinate actuation.

Animators tout the benefits of sketching to develop believable motion parameters where a sketch's subordinated detail presents opportunities to zoom in, problem-solve, then back out — a methodology not lost on haptic designers [58]. Recently developed haptic design tools include Macaron [79], which borrows the animation tenet of keyframes, directly manipulating the vibrotactile sensation to match notable events (or frames) in an analogous animation.

Alternatively, an animator can define a model's movement, e.g., with keyframing. Both techniques can impart expressive or biological-appearing qualities, algorithmically (perhaps with limited quality), or manually (laboriously and with skill). The robot can be triggered to follow the path (pre-computed or generated on-the-fly) by a pre-defined command with a deterministic outcome.

Expressive motion can derive from other sources. "Programming by demonstration" records manually actuated robot motion [21]; actor input can be employed in this way. Hoffman and Ju suggest an iterative approach that integrates robot physical design with 3D modelling for performative robots [39]; Croft and Moon mimic human hesitation behaviours on a robot [54]. Takayama applied traditional animation techniques (such as easing in/out) and tested user perceptions of robot behaviour in a video-based simulated environment [87]. Some generative techniques for affect exist as well: adding Perlin noise to robot poses can increase user recognition rate of displayed emotion [8].

In Voodle, we responded to individuals' natural vocal expressions by providing a novel, direct input mechanism; effectively producing *commands* that modify a pre-determined motion.

2.6 Voice Interaction

2.6.1 Alignment

Alignment happens when behaviours become synchronized in some way. Think about when you are walking with someone, and your strides start to match up. Or when you accidentally start to mimic someone with an accent. This can often be embarrassing, but don't worry—matching speaking patterns is a natural way that humans build trust and ease communication.

This fundamental component of natural human communication is called *alignment*, which occurs when people mimic one another's communicative patterns [33]. Phonetic *convergence* refers specifically to alignment of speakers' phonetic patterning [59], and other studies show that people similarly coordinate speech rhythm, body language and breathing pattern [46, 53, 90]. Similarly, mimicry positively impacts affiliation and likability [46]. Alignment extends to human-computer conversations: people adjust language to their expectations of how a system works [60].

A believable human-robot conversation must likewise see the robot align its communication style, at some level, to the human partner's. Previous work with virtual avatars exploited such linguistic and physical alignment behavior for more naturalistic virtual conversation agents [5, 35, 49]; Hoffman has explored human-robot alignment by utilizing computer vision techniques within performative contexts [38]. Here, we use iconic features of the speech signal to achieve the illusion of alignment, making for more believable interaction.

2.6.2 Iconicity

Speech meaning comes from the semantics of words and phrases, utterance context, the sounds used to construct the words, prosody (tone and rhythm), and accompanying gestures. In the Sausserian tradition, linguistic meaning is an arbitrary relationship between the signifier (a sound pattern) and the signified (a concept) [29]. In this interpretation (symbolic speech), signifier *form* has little relation to its *meaning*. For example, the English word "cat" and its Japanese equivalent ("neko") sound very different, suggesting that the mapping from 'cat' the sound and cat the concept is arbitrary.

The notion of *iconicity* in language is when the form of a word and its meaning are non-arbitrary [61–63]: the word sounds like the thing it represents. For example, the English word for a cat meowing (*"meow"*) sounds very similar to the Mandarin word for a cat meowing (*"miāo"*). Iconic vocalizations are also commonly used to express psychological states (*"ugh"*), or physical phenomena like motion (*"zoom"*) [62].

Iconic vocalizations carry emotional content. Banse and Scherer found that iconic voicing excels in communicating psychological phenomena such as emotional states [6]; Rummer et al demonstrated a relationship between positive emotions and /i/ (the 'ee' sound in 'coffee'), and negative emotions with /o/ (approximately an 'uhh' sound) [70].

Iconic vocalizations are effective for describing physical phenomena and motion.

With physical tools including haptic interfaces, users often opt to use iconic vocalizations to describe tactile sensations [77, 91], and to ground and communicate design intention [4, 14]. Individuals link vocalization features motion patterns with some consistency. Shintel et al saw speakers using high- and low-pitched vocalizations to describe up and down motion respectively. Syllable rate is also a major indicator of visual speed [84]. Voodle uses a similar cross-modal mapping between iconic speech and motion, with upward pitch mapping to upward motion, and time-varying vocal amplitude as a proxy for syllable rate.

Iconicity is an alternative or complementary input mechanism to speech recognition. Previous efforts use sound input to control an interface [34, 42], enhance accessibility of computer systems [10], or as intuitive input for artistic expression [22, 41]. Voice Augmented Manipulation augments users' touch input with voice, e.g., as a modal modifier key [74]. Iconic vocalizations has been explicitly modeled for robots: Breazaeal and Aryananda used prosodic speech features to recognize affective intent, e.g., praise, prohibition, and soothing [13].

In Voodle, we convert vocal features to affective motion rather than categorizing speech.

By utilizing speech form as a basis for controlling robot movement, Voodle can display emotional behaviour without explicit symbolic representation of emotional states. This approach is computationally inexpensive.

2.7 Believability

Believability is a core concept in human-robot interaction design; it is often used, variably defined, and poorly understood. There is a sense in which everyone 'knows' what it means to be believable, but cannot satisfactorily define it. Rose et al [69] attempt to create a quantifiable framework wherein the different senses of *believable* are expressed and interrelated in predicate logic. Roughly, the senses they propose are:

sense 1 A person believes that a robot is capable of an action in a given environment.

sense 2 A robot invokes an involuntary or pre-cognitive reaction in a person similar to that which a non-robot might invoke.

sense 3 A person recognizes the action that the robot performs.

sense 4 A person ascribes mental states to the robot.

Sense 4 is the strongest sense of believability (wherein a person creates a *theory of mind* of the robot) and Rose et al argue that it logically entails the other senses of the word. This partially works: you necessarily have to believe a robot is capable of an action to recognize that action (**sense 3** $\rightarrow$ **sense 1**) and you have to recognize an action to ascribe a mental state to the robot (**sense 4** $\rightarrow$ **sense 3**). But (**sense 4** $\rightarrow$ **sense 2**) is much weaker: you may not need to be involuntarily affected by a robot to ascribe it a mental state[5].

[5]Rose et al recognize this weakness and challenge the reader to find a counterexample. However, it seems that this relationship cannot be directly proved, as **sense 2** needs to be determined *empirically and by analogy*, i.e., you would have to test that there is some involuntary reaction produced by the robot that is not significantly different the reaction produced by some other non-robot entity. Further, the connection to **sense 4** would have to be empirically shown, which, at best, would be correlational. This is not to say that **sense 2** has no use, but it shows that this framework is less logical than heuristic.

This framework is more useful as a set of heuristics than the hoped-for baseline for quantifying believability. **Sense 4** is the closest to the way in which we use believability here.

The notion of believability we use is much more in line with Bates' definition in *The Role of Emotion in Believable Agents* [7]:

> "[A believable character] does not mean an honest or reliable character, but one that provides the illusion of life and thus permits the audience's suspension of disbelief." (Bates 1994)

Key to our understanding of believability is the connection to *narrative frame* and *alignment*. Our understanding of narrative is situated within the constructivist epistemological tradition, which posits that the way in which we construct knowledge about the world is both *subjective* and *negotiated* with the external world. Critically, this is an active and relational understanding of knowledge; we extend this to the notion of believability by saying that an interactor must actively produce the sense of believability in conjunction with the robot. By suspending their disbelief, an interactor creates a narrative for and with the robot, despite their full awareness that the robot is a machine. Paradoxically, this notion of believability has very little to do with an actual belief that the robot is alive; instead, to evaluate how believable the robot is, we would ask whether the narrative is maintained through continued action by both the robot and the interactor that is consistent with that narrative. The question shifts from "do you believe that this robot is alive?" to "are you willing to act as if this robot is alive?"

Connecting this to the concept of alignment and ascribing mental states to the robot, we could say that an interactor has developed a theory of mind for the robot if they are aligned with the robot and if they behave as if the robot has internal thoughts, feelings, or motivations. Again, this can be parallel to their belief that the robot is a machine; rather than being a state of cognitive dissonance, when aligned, the interactor *acts with* the robot to produce a shared narrative context.

In this thesis, we explore how this sense of believability is created, maintained, and lost. We further explore how a robot can support believable interactions using computational techniques such as machine learning, signal processing, and custom software design tools. We analyze the features of the signals that produce robot behaviours and correlate with participant-evaluated emotional interpretations of those behaviours. We posit that believable and emotional robot behaviours can be better understood in terms of complexity measures such as MSE.

2.8 Gestural and Emotional Touch Sensing

A critical component of touch interaction is *sensing* as well as display. Although this thesis is focused on the display of emotional robot behaviours, I have been heavily involved in efforts to sense human emotion through touch gestures by (1) designing, training, and analyzing machine learning models; (2) designing, constructing, and testing custom fabric touch sensors; (3) developing experimental paradigms that support valid emotional interaction; (4) prototyping a simple interactive system to demonstrate mapping from recognized gestures to designed behaviours. This work is reported on here because the machine learning approach to signal processing heavily informed our work on behaviour display, including both generation and analysis.

A first assumption in detecting human emotion through touch interaction is that being able to classify social touch gestures would be beneficial. The intuition is that if you can detect a *pat*, a *happy pat* might be easier to differentiate from a *angry pat*. Using a custom-built touch sensor, in Cang et al [19][6], we explored the detection of social touch gestures using machine learning techniques. Participants touched the CuddleBot given gesture words (pat, stroke, etc.) while touch data was recorded (pressure and location). Using Weka (a common machine learning suite[7]), gestures were detected by breaking the touch data stream into windows and calculating statistical features on the pressure and location data (such as maximum pressure, variance of movement in the X-direction, etc.). We found that we were able to classify touch gestures with over 80 per cent accuracy, depending on the condition. Gesture classification was performed with a random forest using 20-fold cross-validation.

Given our success in touch gesture recognition, we decided to apply the same machine learning techniques to emotional conditions (manuscript in preparation[8],

[6]On which I am a second author, having contributed to study execution, writing, and analysis.
[7]http://www.cs.waikato.ac.nz/ml/weka/
[8]On which I am also a second author, having contributed to study design, execution, analysis, and

also reported on in Cang 2016 [17]). Participants were asked to tell an emotional story while touching the robot. We collected touch, biometric, and eye-tracking data. Here, our success was varied. Given system knowledge of an individual (i.e., using 20-fold cross-validation for model validation), we were able to accurately classify emotional conditions between 70 per cent (touch only) and 99 per cent (integrating touch, biometric, and eye-tracking data). However, given no system knowledge of an individual (i.e., leave-one-out), we were unable to perform better than chance. This lead us to the conclusion that, if it were possible to detect emotional touches, we would have to create an individualized experimental paradigm and machine learning model.

Ongoing work in the SPIN Lab is focused on refining our experimental paradigm and machine learning/signal processing approach to emotion classification.

writing.

Chapter 3

Sketching Robot Bodies

The CuddleBits stem from an attempt to answer design question about simplicity: how expressive can you be with only one degree-of-freedom (DOF)? Given the complexity of rendering emotion on the CuddleBot, the CuddleBits allowed an approach wherein we could decompose the multi-DOF robot into many 1-DOF robots. By studying each expressive DOF independently, we can gain a deeper understanding of how emotional expression works in that DOF, then eventually recompose the DOFs into new multi-DOF robots. For this work, we look primarily at 1-DOF breathing behaviours, explore 1-DOF spine motion, and give an example of how 1-DOF spine and breathing can be recomposed into a 2-DOF robot. To enable rapid iteration, we developed (1) a low-fidelity prototyping method (similar to the design methods one would use to create a puppet); and (2) a design method using *design systems* to enable higher-fidelity iteration. The four CuddleBit design systems that we developed are outlined below. The two major design systems that were studied here are the RibBit and the FlexiBit (see Figure 3.1). The RibBit looks like a rib cage, if you imagine that there are hinges along its spine. The robot expands by the ribs moving outwards. The FlexiBit looks like an orange carefully peeled by slicing into equal sections, if the orange was then taken out and the slices were reattached at the top and bottom. The FlexiBit contracts by pulling the top and bottom together, as if you were squishing the orange peel.

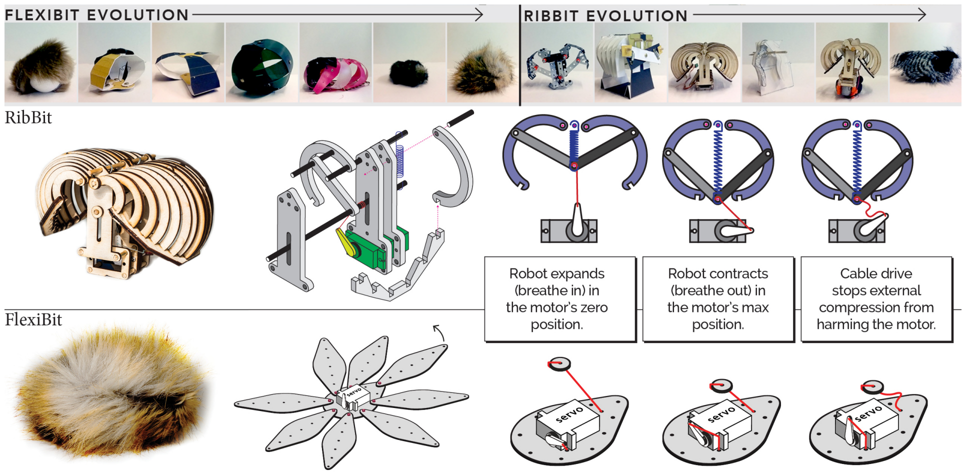


Figure 3.1: Top: the evolution of the RibBit and FlexiBit from low-fidelity paper (and plastic) prototypes to study-ready medium-fidelity prototypes. Bottom: a diagram of the actuation mechanisms of the RibBit and FlexiBit. The RibBit looks like a rib cage, if you imagine that there are hinges along its spine. The robot expands by the ribs moving outwards as figured above. The FlexiBit looks like an orange carefully peeled by slicing into equal sections, if the orange was then taken out and the slices were reattached at the bottom and top.

3.1 Paper Prototyping for Robotics

Paper prototyping is used widely in HCI, where low-fidelity prototypes are made out of simple, often physical media such as paper to rapidly explore and evaluate the design space of an interface. This is similar to a sculptor or designer's use of *maquettes*, where fast, small models of a larger or more complex project are made as studies to begin to sharpen the design intuition about a piece. For example, in theatrical set design, after the initial sketches are made of a potential set, a small scale model is often created to work out design problems on cheap, easily-modifiable media. With a physical model, members of the theatrical production team are able to evaluate the design with more confidence than if the designs had been just on paper, since you can use your spatial intuition to reason about the placement of objects, lighting, etc.

Each CuddleBit design series began with many small, rapidly-produced low-

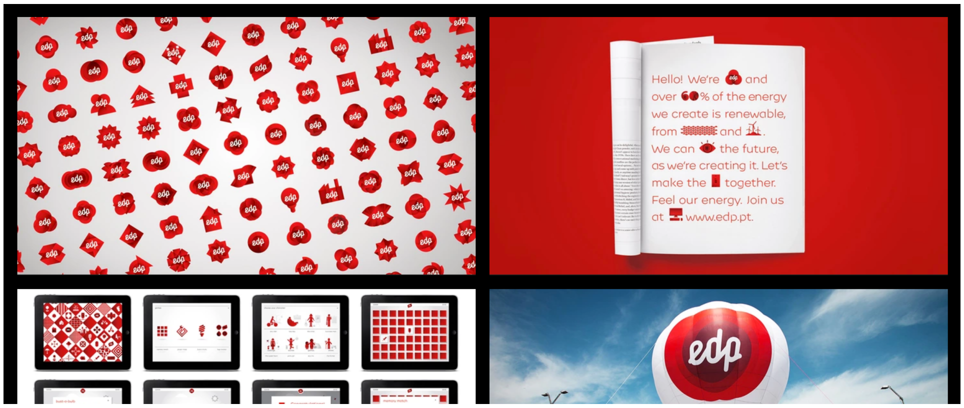


Figure 3.2: Screenshot from Sagmeister and Walsh's website. Notice how a wide number of products can be generated by creating a design system, i.e., an aesthetic, colour palette, and canonical shapes.

fidelity prototypes, often implemented in paper and glue. This allowed for a rapid exploration of actuation mechanisms and form factors with very little cost in terms of time or materials: each prototype cost cents, and could be designed, cut out, and tested in less than a few hours.

The approach of making many incremental designs in parallel is known as *design by prototyping*, and confers the major advantage of having finished prototypes at the end of each design cycle. For the CuddleBits, this meant that they could be used in research very early on in the design process compared to if they had been developed with a traditional engineering design approach, where requirements are meticulously defined ahead of time. In this case, the primary requirement was emotional expressivity; all other requirements were secondary.

3.2 Design Systems

Once the low-fidelity prototyping phase starts to produce converging designs, a higher-level design approach naturally emerges. This happens when commonalities can be identified between designs, including canonical shapes, parts, and actuation methods. Similar to cross-cutting approaches in programming, faster and more powerful designs are enabled when parts can be abstracted and reused. This is the idea behind a *design system*, where a single design effort can produce many finished products.

A *design system* is best explained with an analogy to brand identities. We'll take the famous New York design studio *Sagmeister and Walsh* as an example. They were recently commissioned to create a brand identity for the Portugeuse energy company EDP. *Sagmeister and Walsh* developed a design system based around a common font, colour, core group of shapes, and visual aesthetic. Rather than designing a single one-off logo, business card, or website, the design system could produce many new designs in a variety of visual media with very low impact per new design. By doing the design work upfront, a designer tasked with creating a new product would not have to waste time choosing colors, fonts, or even shapes again, but could build off of the many examples and templates already produced for them.

The CuddleBit design systems take this same approach to robot design. Once working out the fundamental actuation principles, body parts, and assembly methods, new designs could be created that varied in size and shape with relatively little ease. New explorations of expressive capacity could be developed by creating new prototypes within less than a day. Putting the design effort up front allows for fast, dynamic explorations of the design space with low cost[1].

3.2.1 Considerations

Robust compliance and handling affordances: Our prototypes had to withstand the pressure of a human hand, but we found that the structure's basic pliability and apparent fragility also directed its handler's approach: pliability afforded rough play (squishing, hitting, throwing) while rigidity incited gentleness (holding, cupping, stroking).

Touch: Prototypes needed to convey some kind of biological behaviour; thus, the materials needed to afford playfulness and liveliness.

Believability: We drew on caricature and internal consistency for believability; we took cues from the natural world but situated the CuddleBits in their own genre rather than mimicking a real animal. RibBit's inner mechanics are evident, allowing natural material affordances but cueing user expectations; FlexiBit's soft fur

[1]Full explanation of the RibBit and FlexiBit design systems—including assembly instructions—available in the Appendix.

elicited stroking and its limbless form suggested no locomotor ability.

Backend: The Bits are powered and controlled by an Arduino Uno, from a NodeJS server (nodejs.org/en) using the Johnny-Five JS robotics framework (johnny-five.io). Javascript construction facilitates connection to front-end applications and widely-available web frameworks. Using a single language for front- and back-end facilitates seamless development; using web technologies allows for transparency in widget design (i.e., using the browser 'inspect' tool makes for faster debugging than with Java UI packages). As such, rapidly iterating on control mechanisms was also fast.

Extendable and Open design: The CuddleBits were designed to be easily extended with little effort or expertise. Full CuddleBit source documents are shared: pattern files, code and a manual for assembly and extension[2].

3.2.2 CuddleBit Visual Design Systems

Modifiable design systems make varying robot shape quick, from under 2 hours up to 2 days. For each family, we produced many models (Figure 3.1) and chose the most visually and haptically salient to evaluate. Each pattern includes both the atomic units of construction and the narrative sense conveyed by the aesthetic and material presentation, detailed below.

The FlexiBit: Like a simple sewn sphere, Flexi's ribs are plastic slices fixed to a base like petals of a flower. These are generated by adjusting a stencil, printing and cutting the pattern from plastic sheets with a knife, and joining them with machine screws. Slices are scalable for smaller or larger Bits; to adjust shape, only some slices and/or the base are varied. Plastic flexibility, volume, and curvature provide passive compliance and natural feel under a faux fur cover. It is often compared to a Tribble (a fuzzy alien species from the Star Trek universe): the plastic frame evokes a compliant torso.

The RibBit: A wooden ribcage on a stand, its rigid actuation gets compliance from internal springs. In counterpoint to Flexi, mechanics are fully exposed, with no attempt at material realism. It comes alive only with suggestive movement. Each rib is laser cut from an easily modifiable digital pattern, and assembled by

[2]www.cs.ubc.ca/labs/spin/cuddlebits: pattern files, code and a DIY manual for assembly and extension.

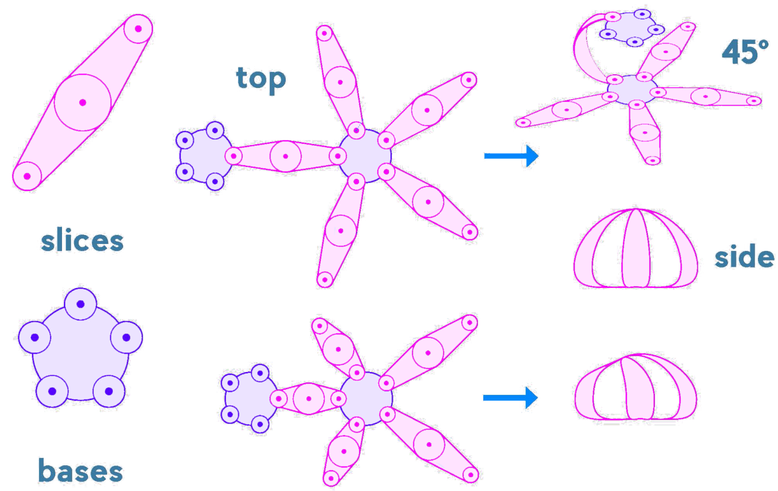


Figure 3.3: FlexiBit design system. Using two canonical shapes, the slice and the base, the FlexiBit's shape and size can be quickly and easily varied.

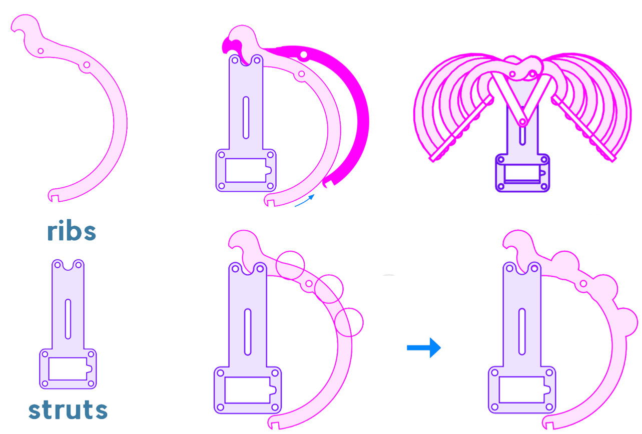


Figure 3.4: RibBit design system. Using Adobe Illustrator, it is easy to modify the texture, shape, size, and number of the ribs, making iteration quick and easy.

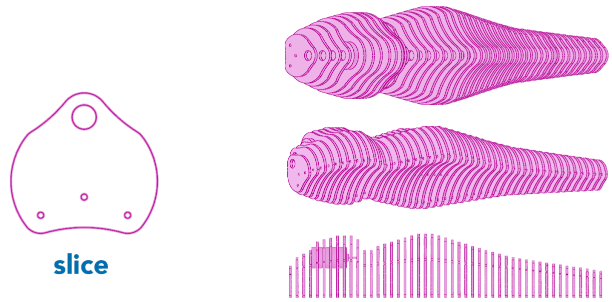


Figure 3.5: SpineBit design system. The whole robot is built off of configurations of a single slice (shown) which is defined by parameterized curves. By defining key slices and interpolating between them, a new robot shape can be produced.

wood-gluing parts together, with BBQ skewers as pins and rods. Further versions of the RibBit include fur and ridges that form an inflexible ‘spine’; these explorations were prototyped in low cost materials (like paper) first.

RibBit has a naturalistic skeletal aesthetic due to its wooden construction. Fitting comfortably in the contours of a hand, the structural rigidity provides notable haptic feedback even when covered with fur.

The SpineBit: Made from many ‘slices’ of wood strung together by elastic bands and seperated by small spacers, the SpineBit passively curls around objects and body parts to give the effect of ‘hugging’. Similar to the RibBit, the SpineBit has a hard and skeletal aesthetic, but the use of elastics gives the shape pliability.

3.3 The ComboBit

The ComboBit combines the spine actuation of the SpineBit with the flexible breathing concept of the FlexiBit. This robot is the first attempt at extending our prototyping process from one- to two-DOF, and is a work in progress.

Before settling on the FlexiBit as the principle ’Bit to integrate with the SpineBit, a number of RibBit-and-SpineBit configurations were attempted. The limiting factor was the mechanical dependence of the spine and the ribs. To support bending, different ribs need to be actuated at different rates, making the coupled RibBit rib

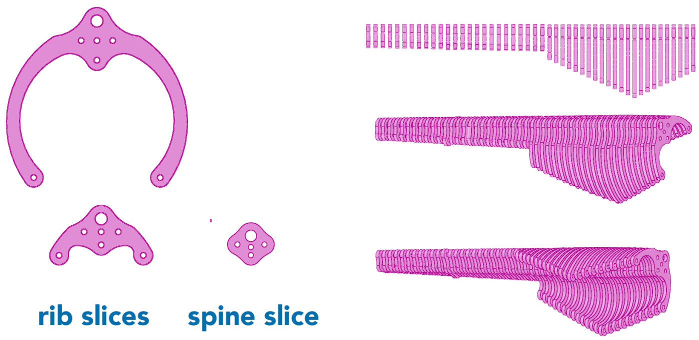


Figure 3.6: ComboBit design system. The robot is built off of configurations of a two slices (two configurations of the same 'rib' slice shown left; one configuration of the 'spine' slice shown right) which is defined by parameterized curves. By defining key slices and interpolating between them, a new robot shape can be produced.

cage design insufficient. An attempt was made to create individually-actuated ribs. However, the mechanical complexity created high friction and high chance for mechanical failure. Instead, a FlexiBit-inspired design—where the elasticity of the body material allows for both compression and expansion—affords actuation in multiple directions.

Although the ComboBit is a blend of previously-designed robots, we produced low-fidelity prototypes at all points during the design process. Hand-modifiable materials such as paper afforded designs to be quickly and cheaply tested on the fly. Rather than being a step backwards, returning to low-fidelity prototypes after having created higher-fidelity prototypes focused design efforts on new design ideas.

Both the ComboBit and the SpineBit are designed using a Solidworks design table, which allows for designing with equations, relations, and variables. Since Solidworks is a constraint-based parametric design suite, the slices that make up the 'Bits designed with a canonical set of shapes and relationships where dimensions are variable and controlled via the design table (see Figure 3.7). By determining the dimensions of 'key' slices, dimensions to vary, and interpolation type (linear, polynomial, spline, etc.), it is possible to fill in configurations between the

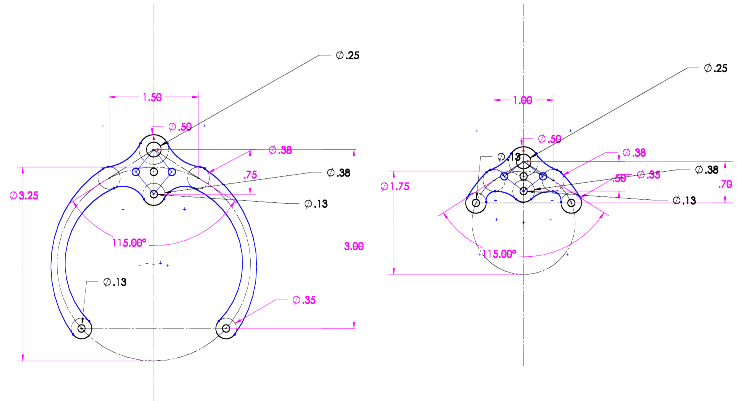


Figure 3.7: Exampe of two slices created by design tables. Notice that the defined curves are the same for both: the same number of circles and curves in the same relative placements, with the same relations (i.e., some point X is coincident with some circle K) with the different values for the same dimensions. The values of the dimensions are set in an Excel table, which is used by Solidworks to produce different configurations of the same shapes and relations.

key slices to create a smooth, continuous body. Effectively, this allows for rapid iteration in the shape and size of the robot as long as the actuation mechanisms are kept unchanged.

3.4 Moving from 1-DOF to 2-DOF

Work is ongoing to fully integrate spine and breathing motion into a 2-DOF robot with the ComboBit. Lessons learned from the design process with the ComboBit are instructive for determining the extent to which our chosen deconstructive design process will be successful. Here summarized are insights and questions from the early attempts at the ComboBit.

Starting from scratch: Since the designer has already spent a lot of time becoming familiar with the materials, actuation principles, and shapes of each design system during previous iterations, they do not start from ‘scratch’ each time. Even early lo-fidelity prototypes start sophisticated. This comes with a tradeoff of full exploration of the design space. Since the designer has design solutions in mind already, it is less likely that they will diversify their approach. To overcome this, it is necessary take alternative approaches to determine viability. The extent to which the designer takes the project of multiple designs seriously is what determines the extent to which they will have to start again from nothing.

Tacked-on vs. fully integrated: Certain DOFs are more difficult to 'tack on' than others. For example, one can imagine fairly easily adding a rumble motor to any CuddleBit design with relatively little design work. Similarly, with the ComboBit, one could imagine a design solution where the parts of the robot that bend (spine) and breathe are not directly connected. However, the ComboBit attempts a 'full' integration where the breathing and spine part are mechanically dependent. This greatly increases the design time needed, but also increases the extent to which the robot 'feels real', as the mechanically dependent parts move together (e.g., like a spine and rib cage).

Dependence on prototyping materials and processes: Since the robots are being designed iteratively with the goal of having working prototypes finished at the end of each stage, the designs are highly dependent on the prototyping material and processes used. For example, since lasercutting and 3D printing machines are readily available, the robots have been designed to leverage those technologies and use materials that they support. This has the advantage of being highly accessible, i.e., anyone with access to rapid prototyping machines can easily create their own CuddleBits. However, as the robots become more complex, this design approach limits the integration of more sophisticated materials and mechanisms. Neither is better; the success of the project is dependent on how the robots are intended to be used.

Committing to the ease of future iteration: With more moving parts, designing for future iteration is more difficult with 2-DOF relative to 1-DOF. However, it is still possible with some high-level planning of the dimensions you wish to control. With a SolidWorks design table, you are able to choose the dimensions over which you vary: once the base shapes and relations are defined, any variation that directly scales any of the dimensions is just a matter of inputting new numbers. Even with a design table, this ability to vary takes design time, since relations and shapes need to be tweaked to keep the model well defined.

Iterations are relative to the design system: The new robot designs that are enabled by the design systems are necessarily constrained by the design systems as well. The space of iteration is somewhat predefined: the dimensions along which iteration is to be performed has to be explicitly designed for. Therefore, some changes are easy (i.e., changing scale, adding/removing slices/ribs, textural changes), but some changes require defining a whole new design system, in which case, the designer is doing more work than simply making a one-off design. The design system approach is only appropriate when many smaller variations of robot are required.

3.5 Conclusions

The 1-DOF CuddleBits are easy to build, modify, and iterate on. They show good potential for integrating into 2-DOF robots. The CuddleBits were built using a sketch-and-refine paradigm, where they were first built with hand-modifiable lo-fidelity materials such as paper and light plastic, and a wide variety of form factors and actuation principles were explored. Keeping the principle of *design for re-design* in mind, the CuddleBits were then further developed as *design systems* that allowed for many configurations of a robot to be built using the same base design. This approach allowed for rapid exploration of the design space where expressivity was the key metric for valuation.

Chapter 4

Generating Robot Behaviours

The simplicity of the CuddleBits cannot be overstated. One reviewer—in a positive review—called them "painfully simple," and I heartily agree. That is the fundamental power of the work. We were able to display a surprisingly wide range of emotions using very simple machines. The breathing-type CuddleBits we study here (the RibBit and the FlexiBit) are only able to expand and contract a portion of their bodies[1]. We used internally developed haptic design tools to quickly sketch and refine behaviours: (1) MacaronBit, a keyframe-based vibrotactile effect editor and (2) Voodle, a robot behaviour sketching tool.

When designing physical objects, a designer will typically start by roughly *sketching* the object, then iteratively *refining* their design. *Sketching* is improvisational and rough, and allows a designer to test many different ideas by quickly creating and evaluating their designs on the fly. When *refining*, the design is much more static, and changes are incremental. After attempting a number of sketching and refining tools for robot behaviour design, our group converged on Voodle/-MacaronBit. Here, we discuss the development and evaluation of our design tools wherein we (a) conducted a pilot study to determine which aspects of vocal input were most salient for robot motion; (b) conducted a behaviour design study where participants were asked to create emotional designs using Voodle and MacaronBit.

[1] They basically just wiggle.

4.1 Early tool attempts

While extending Macaron for robot behaviour design seems natural in retrospect, at the time it was not obvious to repurpose a vibrotactile design tool for 1-DOF robot motions. A keyframe editor balances the concerns of precision and refinement with the easy sketching abilities afforded through direct manipulation; Macaron provides focused control over a short window (in the order of seconds).

Our group made many attempts at behaviour design/puppeting tools and techniques before MacaronBit and Voodle, including a free-hand vector drawing tool based on paperjs.org, a browser-based timeline editor [2]; and for direct position control, a force-feedback knob [83]. The timeline editor was unintuitive; the drawing tool and force-feedback knob required too much fine control for large movements, and did not provide enough control for small movements (i.e., entire behaviour would have to be re-drawn per sketch; human hands cannot move fast enough to express fine motions such as fluttering). MacaronBit allowed for precise control of both large and fine motions, and Voodle allowed for quick sketching and iteration since it gave immediate continuous feedback.

In later work, we attempted to use genetic algorithms with complexity measures as utility functions to generate behaviours at a variety of valence levels. However, even for relatively short behaviours, the process of generation was extremely long. As such, we returned to using Voodle and MacaronBit to sketch and refine behaviours.

4.2 MacaronBit: A Keyframe Editor for Robot Motion

Macaron is a open-source web-based keyframe editor for designing vibrotactile sensations, using amplitude and frequency of a vibration [79]. As our first pass at designing a robot behaviour editor, we extended Macaron to robot position control, calling the result MacaronBit (Figure 4.1).

In developing MacaronBit, we started with a pure sine-wave and adjusted its parameters: frequency, amplitude, bias, and amplitude/frequency variability. Its support of immediate playback, key-framing (parameter interpolation between key points), waveform generation, and click-and-drag editing sped up iteration. For participant-designed behaviours, we switched to direct position control, where de-

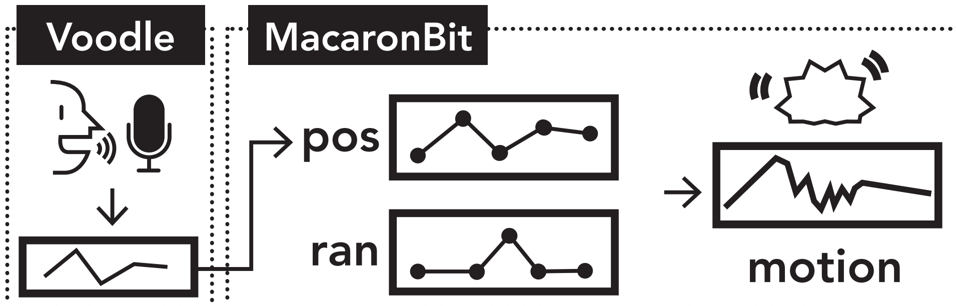


Figure 4.1: For behaviour generation, users had two design tools to help create behaviours: Voodle and MacaronBit. With Voodle, users could optionally sketch a robot motion using their voice. Their vocal input was imported into MacaronBit as raw position keyframes (shown as 'pos' above). Users could modify the waveform by manipulating other parameters, specifically randomness (shown as 'ran'), max and min position. MacaronBit includes standard keyframe editing functions.

sign parameters were position, randomness, and max and min position (Figure 4.1).

4.3 Voodle: Vocal Doodling for Affective Robot Motion

Interactive agents are more compelling when they are believable: giving the illusion of life and facilitating suspension of disbelief [7]. When users believe that an agent has a 'spark of life', they can be more immersed, emotionally invested, and aligned with the agent system.

However, creating believable agents is hard. Animators and roboticists are highly trained, use cutting-edge modeling tools, and have to balance making their animations too real (and becoming uncanny) and not real enough (thereby not being understood). Yet actors and performers *improvise* believable characters. While this may require skill, effort and a specific state of mind, the effort is applied directly, not through a computer keyboard or by writing code. Can performance be leveraged to improvise believable robot motion?

We focus here on voice, for two crucial qualities. First, voice naturally expresses emotion; meaning is conveyed through the *form* of speech (e.g., prosody) as well as utterance semantics. A dog can thus take direction from its owner's

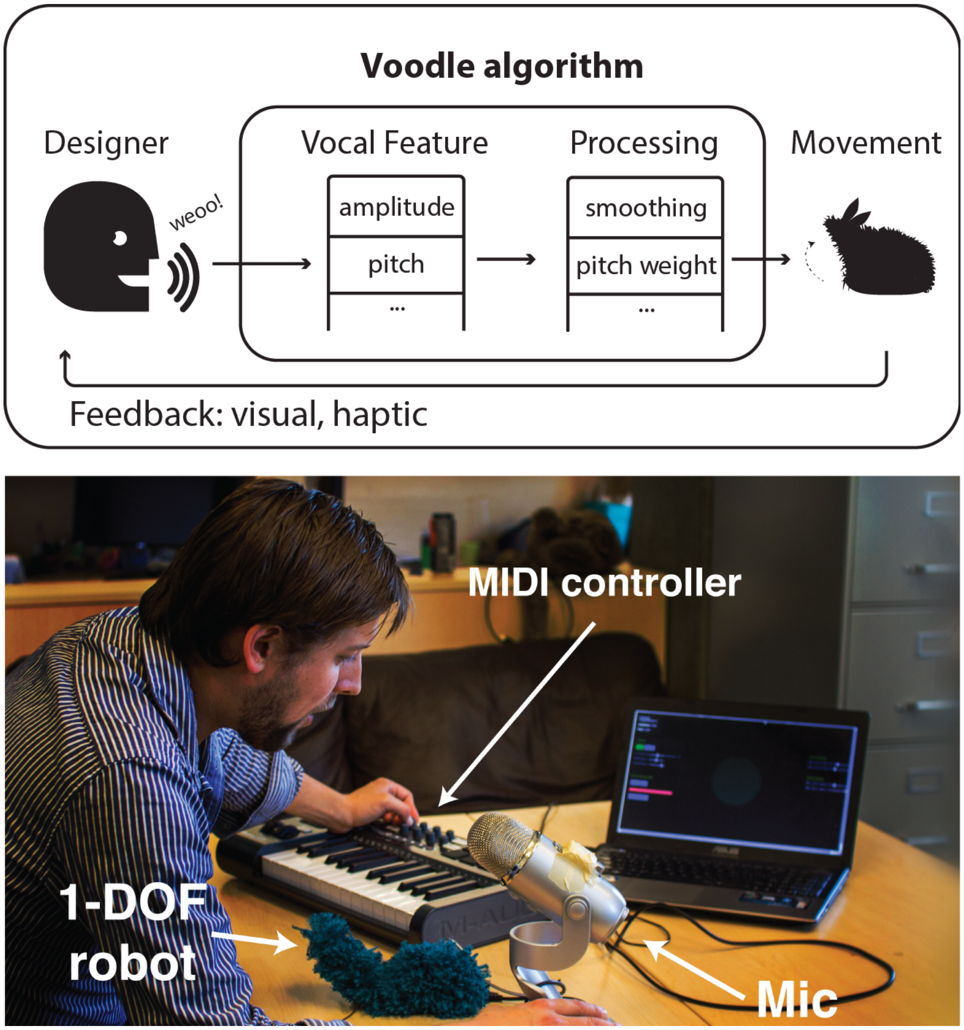


Figure 4.2: Voodle (vocal doodling) uses vocal performance to create believable, affective robot behaviour. Features like tone and rhythm are translated to influence a robot's movement.

tone, timing and loudness as well as her words. Secondly, iconic sounds, found in onomatopoeic words like *"boom"*, *"woof"*, or *"ding,"* can capture hard-to-express ideas like emotion (ugh) or movement (zoom).

We posit that an *iconic vocabulary* could be the basis of a rich, naturalistic, and improvisational platform to interactively design behaviors for physical, affective systems.

To assess this proposition, we built the Voodle system ('vocal doodling'), which derives believable motion from iconic vocalizations[2]. Specifically, we used computationally low-cost methods such as real-time amplitude and pitch analysis of vocal performances to immediately generate motion on a 1-degree-of-freedom (DOF) robot, allowing the performer to evolve and experiment as he seeks a particular behavior. Voodle was developed as a *design tool*. That is, we intended that roboticists would perform vocalizations that move a robot as a way to design its behaviour. However, along the way we also discovered its promise as an *interaction technique*: an expressive input that end-users can employ to elicit lifelike motion as they interact with the robot.

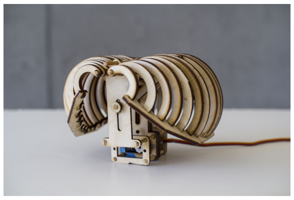
A 1-DOF robot can be expressive yet relatively easy to implement and control, and offers insight into motion for more complex robots. Our final Voodle system mapped increased pitch and amplitude to CuddleBit height. We first describe a pilot study to gather requirements for a working system, then report implementation details.

4.3.1 Pilot Study: Gathering Requirements

We conducted a pilot to inform an initial Voodle implementation based on the RibBit (Figure 4.3). Like most of the Bits, the RibBit moves its "ribs" in and out with a breathing-like motion. To identify and prioritize features, we captured vocalizations people use to describe robot behaviours, characterized how people mapped sounds to robot movements, and identified key vocal and system features for implementation.

We recruited five participants (aged 20-26, 2 female) from a university population, reimbursed $10 for a 1-hour session. All were fluent in English (four native speakers, one native Russian speaker; four multilingual) with varied artistic and performance experience, e.g., acting, illustrating, music.

[2]We use "Voodle" to refer to our implemented system, "voodling" to the act of using iconic vocalizations as an input modality with an interactive system, and "voodles" for specific vocalizations.



(a) RibBit

(b) FlexiBit

(c) FlappyBit

(d) VibroBit

(e) BrightBit

Figure 4.3: The 1-DOF CuddleBit robots used in the Voodle co-design study: (a) RibBit: A CuddleBit that looks like a set of ribs; (b) FlexiBit: A Bit whose stomach ”breaths” via a servo; (c) FlappyBit: A Bit with an appendage that flaps up and down via a servo; (d) VibroBit: A Bit that vibrates via an eccentric mass motor; (e) BrightBit: A Bit whose eyes light up via an LED.

Methods

After an icebreaker activity (tongue-twisters and improv game), participants completed a vocal imitation task, then a vocal improvisation task.

Imitation task: Participants observed and optionally used their hands to feel each of 18 movements through the robot, then imitated the behaviour using iconic vocalizations.

Of the 18 robot motions, ten were developed using vibrotactile signals from an existing library that categorizes vibrations based on perceived dimensions such as energy, duration, rhythm, roughness, pleasantness, and urgency [82]. These had been previously chosen for the purpose of expressive vibrotactile display, by two researchers independently selecting exemplary vibrations from the library’s dimensional extremes then iteratively merging their choices [80].

We produced eight more motions by systematically varying sine parameters: fast/slow, large/small, and rough/smooth.

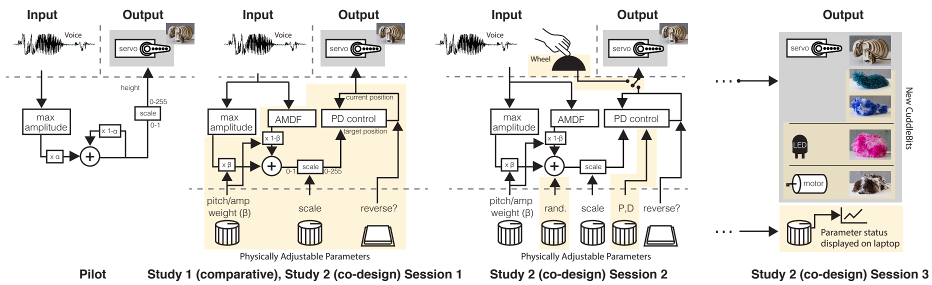


Figure 4.4: The Voodle system implementation, as it evolved during our studies. Additions for each stage are highlighted in yellow. In our final system, incoming vocal input is analyzed for amplitude and fundamental frequency. These signals are normalized between 0 and 1, then averaged, weighted by a "pitch/amp" bias parameter. Randomness is then inserted into the system, which we found increased a sense of agency. Output is smoothed either with a low-pass filter or PD control. Final output can be reversed to accommodate user narratives (i.e., robot is crunching with louder voice vs. robot is expanding) for several different CuddleBits.

Motion durations ranged from 1-13 seconds and were looped.

Improvisation task: Participants manually puppeted the unpowered robot while spontaneously vocalizing their puppetry, while audio and video were recorded.

Analysis: We transcribed vocalizations into the International Phonetic Alphabet (IPA) from the imitation task to capture and prioritize input sounds, observed and reported how people mapped sounds to robot movements, and observed phonological similarity within and between participants.

Table 4.1: Pilot Study: Linguistic features that participants felt corresponded best with robot position in the imitation task. "+" and "-" indicate feature presence or absence. The comparative Study 1 went on to use *pitch* as a primary design element.

Feature	**Feature Description**	**Example Tokens**	**Dominant Participant-Produced Behaviours**
Pitch	Perceived fundamental frequency of the vocalization over time.	*"dum DUM"* [↘dum ↗dum] *"We eEH"* [↘we ↗e] *"mMm"* [↗mː↘m]	Upward movements associated with higher pitches, and downward movements associated with lower pitch; sometimes reversed.
+/- Continuant	Whether or not airflow is fully obstructed in the vocal tract during speech, e.g., , the "f" in "father" vs the "t" in "butter"	*"waywayway"* [weiweiwei] (+continuant) *"dum dum"* [dʌm dʌm] (-continuant)	Continuants are associated with behaviours that begin with gradual and smooth motion, while non-continuants are associated with behaviours with abrupt and jerky motion.
+Strident	When there is a large degree of turbulence and high energy noise caused by an obstruction in vocal tract. Example: the *"sh"* in *"shush"*.	*"tchuh-tchuh"* [t͡ʃʌ.t͡ʃʌ] *"tcheen"* [tʃin]	Rapid movements – e.g., , the Bit moves very quickly between different positions.
+/- Voiced consonants	A consonant is voiced if it's produced while the vocal folds are vibrating.	*"ga"* [ga] (voiced) *"ka"* [ka] (unvoiced)	Voiced consonants were associated with smooth motion, while unvoiced consonants were associated with less smooth motion.

Results

Phonetic Features: Table 4.1 reports typical phonetic features that we observed in the pilot study's imitation task. We transcribed vocalizations into the International Phonetic Alphabet (IPA), then organized them by distinctive phonological features [20]. The most compelling features, based on discriminability on motion and feasibility of implementation, were pitch, continuants, stridents, and voiced consonants.

Metaphors for Sound-to-Behaviour Mappings:

Participants instituted a relationship between pitch, amplitude and height: the higher the robot's ribs, the higher the pitch and amplitude.

There were exceptions to this pattern; for example, one participant saw the robot's downward movement as 'flexing,' and therefore used increased vocal pitch and amplitude to represent its downward movement. Table 4.1 reports contrasting relationships that we observed, with examples.

We saw occasional reversals in participants' mappings between the imitation task and the improvisational task.

One possible cause is the Bit's actuation methods: i.e., , computer-control in imitation, and participant-actuated in improvisation. The only direction to manually actuate the robot is downwards: its default state is an extended position, and the ribs are normally pulled inwards by a servo. Hence, increased physical effort translates to downward movement. So the relationship between pitch and amplitude may be based on how the participant conceptualizes the "direction" yielded by the work.

Individualized language: Each participant seemed to have idiosyncratic sound patterning. For example, some participants used many voiced stops (e.g., *"badum badum"*) in their utterances. Some participants consistently used multiple syllables with many consonants (*"tschugga tschugga"*); others consistently produced simple monosyllabic utterances (*"mmmm"*).

4.3.2 Voodle Implementation

Based on piloting guidance, we created a full Voodle system, seen in Figures 4.4 (system design) and 4.2 (system in use).

We found that fundamental frequency and overall amplitude (easily detected in realtime) could capture a variety of relevant vocalizations, including pitch and +continuant features. To accommodate variety in metaphors (e.g., breathing vs. flexing) and individualized language, we included user-adjustable parameters: motion smoothing, gain, pitch and amplitude weight (where the weight between amplitude and pitch is a linear combination: $output = amp \times amp_{weight} + pitch \times pitch_{weight}$), and the reverse. Priorities for future phonetic features include distinguishing the additional features reported in Table 4.1.

Voodle was implemented in JavaScript: a NodeJS server connected with the RibBit using Johnny-Five and ReactJS [43, 66].

Input audio was analyzed in 1s windows. Amplitude was determined by the maximum value in the window, deemed to be sufficient through piloting. The fundamental frequency was calculated using the AMDF algorithm [85], the best performer in informal piloting. Figure 4.4 shows algorithm evolution. Voodle is open-source, available at https://github.com/ubcspin/Voodle.

4.4 Evaluating Voodle and MacaronBit

To develop a set of behaviours to evaluate for emotion content, we ran a study wherein participants were asked to work with an expert animator to create behaviours given an emotion word.

4.4.1 Methods

We recruited ten participants to design five robot behaviours, each based on an emotion word from the PANAS scale [93]. Three self-identified as singers or actors.

To define the design tasks, participants were assigned one word per affect grid quadrant, chosen randomly without replacement from the five PANAS words for that quadrant; participants selected a fifth word. The words were presented in random order.

Unpleasant Activated	**Pleasant Deactivated**	**Pleasant Activated**	**Unpleasant Deactivated**
stressed† *	relaxed†‡*	excited†‡*	depressed†*
upset‡*	calm‡*	attentive‡*	drowsy‡*
scared‡*	at rest‡*	determined‡*	bored‡*
guilty‡	serene‡	proud‡	dull‡
hostile‡	at ease‡	enthusiastic‡	sluggish‡
nervous‡			droopy‡

Table 4.2: Affect Grid quadrants of PANAS emotion words. † represents words used in CuddleBits Behaviour Generation Study; ‡represents words used in CuddleBits Study 1 (see *Evaluating*); * represents words used in CuddleBits Study 2 (see *Evaluating*).

For each word, the participant was given the option to express the behaviour with Voodle, design it using a traditional keyframe editor, or switch between these as needed.

The keyframe editor, Macaron [78], allows users to specify Bit height (periodic movement amplitude) over time, as well as remix and transform their original animations through copy/pasting, scaling keyframes, and inversion and other functions. Participants could export their voodles as keyframe data for later refinement in Macaron.

During the study, an expert animator (a co-author) was a design assistant, introducing participants to the robot and two tools.

The animator assisted participants in creating compelling designs, offering technical support and guidance as needed, but did not create animations for them. Meanwhile, another researcher acted as an observer, taking notes on tool use and conducting a brief informal exit interview.

Participants could create as many designs for each emotion word as they wanted using any tool at any time until they were satisfied with the result; for example, they might make three designs for “excited” and choose their favourite.

4.4.2 Results

Voodle was used beforehand to sketch behaviours in most cases. When a participant had a clear idea of what the behaviour should look like, both sketching and refining was performed in MacaronBit.

A library of 72 behaviours labelled by emotion word was generated (participants designed multiple behaviours per word); analysis of these behaviours presented in *CuddleBits: Study 2* (*Evaluating* 5) results.

Participants agreed that the robots came to life: *"it shocked me how alive it felt," "it tries to behave like a living thing would."*

Voodling was used by participants to express emotions: *"the things [Voodle]'s listening for is different from the things Siri listens for...it's usually emotional meaning or mental state that's conveyed by [pitch, volume and quality]"*. While 7/10 participants used Voodle, those with performance experience experience used Voodle more. This is may be individual preference: voodling is performance, and tended to be preferred by those comfortable with performing.

Participants generally chose to use Voodle to augment their keyframe-editor work, rather than as a stand-alone tool. Only two (both performers) ever designed with Voodle alone, and only did so for one behaviour design task each.

Voodle was most appropriate for exploring and sketching ideas, not fine-tuned control. When users knew their goal, they moved straight to the keyframe editor: *"it always seemed easier to go to [the keyframe] editor to do what I had in my head than trying to vocalize and create that through voice."*

We found participants had trouble expressing static emotional states (e.g., , *distressed*); these became clearer when contrasted with an opposing emotion. In our next study with Voodle, we changed the task to transitions between emotional states.

Supplementing these observations, we note that a concurrent study (whose focus was on developing and assessing these robots' expressive capacity, and not on input tools) also used these Voodle-generated animations along with others, and confirmed that they covered a large emotional space [16]. Specifically, independent judges consistently assessed Bit animations as well-distributed across the arousal dimension, and somewhat along valence.

We concluded that Voodle had value for sketching expressive robot behaviours, but needed further development.

4.5 Conclusions

In this section, we discussed the development and assessment of our two robot behaviour design tools, Voodle and MacaronBit. The former is used for sketching and the latter for refining robot behaviour designs. Both tools have been used so far for designing 1-DOF robot behaviours; there is evidence that extensions to multiple DOFs are possible, but the transition to multiple DOFs would not be simple. Like many 3D keyframe animation tools, MacaronBit may need a more complex timeline-based approach where multiple DOFs and complex movements are controlled through composing simpler movements (i.e., by joining, nesting, etc.). A vision for Voodle may be to set two keyframe multi-DOF robot positions, then to use voice to interpolate between them. For example, imagine a humanoid robot that is going from crouching to standing.

Chapter 5

Evaluating Robot Behaviours

To determine the ability of the CuddleBits to display a wide range of emotions, we ran two studies where participants rated breathing behaviours in terms of arousal and valence (*CuddleBits* studies 1 and 2). We then ran a six-week co-design study with expert performers to further develop our improvisation tool and study how behaviour design could work when conveying changing emotional behaviours, i.e., from *stressed* to *relaxed* (*Voodle*). Last, we ran a study wherein we explored the relationship between complexity, valence, and narrative context, wherein participants rated robot behaviours for valence and created short stories that explained the robot behaviours (*Complexity*).

5.1 CuddleBits Study 1: Robot Form Factor and Behaviour Display

We evaluated the emotional expression capabilities of our two CuddleBit forms (FlexiBit and RibBit) on eight behaviours representing four emotional states. Specifically, we asked:

RQ 1. Can 1-DOF robot movements be perceived as communicating different valence and arousal states?
Hypothesis: Different levels of arousal will be interpreted more accurately than different levels of valence.

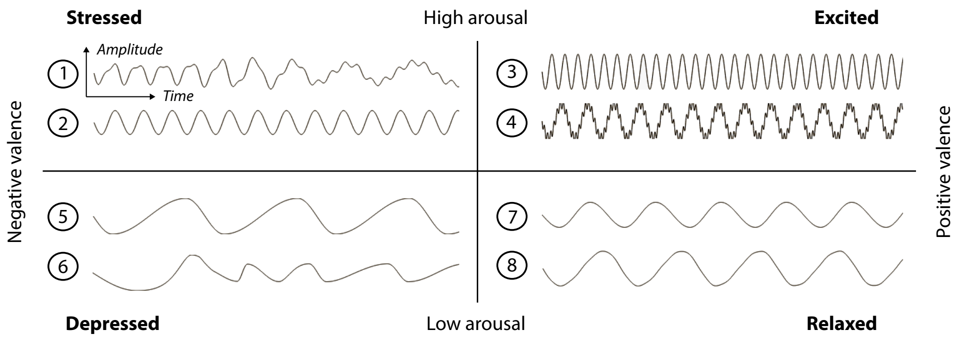


Figure 5.1: Waveforms of Study 1 behaviours as designed by researchers. Each quadrant is represented by a PANAS affect word corresponding to the extremes along (valence, arousal) axes, i.e., *Excited* is high-arousal, positive-valence.

RQ 2. How is interpretation of emotional content influenced by robot materiality, e.g., a soft furry texture?
Hypothesis: FlexiBit's behaviour will be perceived as conveying more positive valence than RibBit's.

5.1.1 CuddleBits Study 1: Methods

Behaviour design: Team members created and agreed upon two breathing behaviours for each quadrant of the affective grid [72]: *Depressed*, *Excited*, *Relaxed*, or *Stressed*, for a total of 8 behaviours (represented as motion waveforms in Figure 5.1). Each emotion word typifies the extreme of its emotion quadrant (i.e., *Stressed* is high-arousal, negative-valence).

Participants: 20 participants, aged 20–40 with cultural backgrounds from North America, Europe, Southeast Asia, Middle East and Africa, were compensated \$5 for 30 minute sessions.

Procedure: Participants were given the task of rating each behaviour on a 5-point semantic differential (-2 Mismatch to $+2$ Match) for two different robots displaying four emotions: *Depressed*, *Excited*, *Relaxed*, or *Stressed*. For instance, for "FlexiBit feels stressed", a participant would play each behaviour and rate how well it matched the robot portraying stress. During playback and rating, participants kept one hand on the robot, and moused with the other; motion was experienced largely haptically. Noise-cancelling headphones played pink noise to mask mechanical noises; instructions were communicated by microphone.

Ratings for each robot were performed separately. Robot block order was counterbalanced, with an enforced 2m rest. For each block, all four emotions were presented on the same screen so participants could compare globally. Behaviours (15s clips) could be played at will during the block. Order of behaviours and emotion was randomised by participant. To reduce cognitive load, participants saw the same behaviour/emotion order for the second block. In total, each participant performed 64 ratings (8 behaviours $\times$ 4 emotions $\times$ 2 robots). Afterwards, a semi-structured interview was conducted.

5.1.2 CuddleBits Study 1: Results

We compared ratings of each pair of behaviours designed for the same emotion word with a pairwise Wilcoxon signed-rank tests with Bonferroni correction (Figure 5.2). Ratings of the two designed behaviours for the same emotion quadrant were not significantly different ($\alpha = .050/8 = .006$; all p's $\geq .059$). Thus, we averaged ratings into four pairs by emotion target (e.g., (1) & (2) in Figure 5.1).

Effect of emotion quadrant on behaviour ratings (significant). Friedman's test on behaviour ratings showed significant differences between behaviours per emotion for both robots (all p's $< .001$). Post hoc analyses using Wilcoxon signed-rank tests were conducted with a Bonferroni correction ($\alpha = .050/6 = .008$) to further analyse the effect of **emotion condition** on *researcher-designed behaviours*:

– **Stressed**, **Excited**, or **Relaxed**: There were significant differences between high and low arousal behaviours (*Stressed-Depressed*, *Stressed-Relaxed*, *Excited-Depressed* and *Excited-Relaxed*, all p's $\leq .002$); but none between behaviours with the same arousal level but different valence content.

Effect of robot on behaviour ratings (not significant). Wilcoxon signed-rank tests with Bonferroni correction showed no statistically significant differences between ratings of emotions displayed on the two distinct robot forms ($\alpha = .050/16 = .003$; all p's $\geq .026$).

Duration (not significant). A two-way (2 robots $\times$ 4 emotions) repeated measures ANOVA showed no significant differences in the time spent on rating behaviours (all p's $\geq$ .079), suggesting each emotion rating was undertaken with similar care.

5.1.3 CuddleBits Study 1: Takeaways

Hypothesis 1: Different levels of arousal are easier to interpret than different levels of valence. – *Supported.*

In general, participants were able to perceive differences in behaviours designed to convey high or low arousal. Speed or frequency was most mentioned for arousal variation: low arousal from low frequency and high arousal from high frequency. Participants found interpreting valence more difficult. Thus, behaviours on this 1-DOF display corroborates earlier findings in regards to both dimensions [27, 65, 97].

We posit that the difficulties in determining valence may be due in part to the restrictive range of behaviours. All designs were based on the perception and imagination of three computer science researchers, which may not be broadly generalizable as effective emotional displays.

Improvement: Behaviours may have more range or discernible valence when sourced from a more diverse group of designers. To increase emotional variance in Study 2, *we recruited participants (N=10), the majority of whom were employed in creative roles to create the behaviours with an expert designer.* Participants were encouraged to puppet robot movements, act out desired movements, and interact with the robot until they were satisfied with the emotional displays.

Hypothesis 2: FlexiBit's behaviour will be perceived as conveying more positive valence than RibBit's. – *Not supported.*

In post-study interviews, participants reported the movement expressed by the two robot forms as sensorially but not necessarily emotionally different. FlexiBit felt nicer to touch, but its motion was less precise. RibBit's movements were interpreted as breathing or a heartbeat despite the exposed inner workings emphasizing the 'machine-ness' of the robot.

Unexpectedly, while participants specified preferences for FlexiBit's fur and

RibBit's motor precision, pairwise comparisons of the same emotions revealed no significant difference between robots. Movement rather than materiality dominated how participants interpreted emotional expression; although visual access to form was restricted during movement, tactility might have modulated perception of, e.g., life-likeness.

Improvement: Whereas robot form factor had little to no influence on emotion recognition results, it did influence how participants perceived the robot. We selected characteristics to emphasize for a second round of robot prototyping, *producing a new robot for Study 2*. We focused on characteristics that participants referenced as salient or pleasing in interviews, such as fur, texture, and body firmness.

Starting from paper prototypes, we iterated on the RibBit form factor to increase haptic salience and to incorporate positive FlexiBit features. After exploring bumps on the ribs, spine configuration, fur textures, and rib count, we converged on a form that had fewer ribs, dense fur, and a prominent spine. This combined the favourite features of the RibBit (crisp motion and haptic feedback) with the FlexiBit's cuddliness. With rapid prototyping methods, each paper/lo-fi sketch could be explored in less than an hour; full new robot prototypes took about two hours to modify design files, half an hour to laser cut, and about two hours to assemble.

5.2 CuddleBits Study 2: Evaluating Behaviours

In a second study, we validated our participant-created behaviour designs and explored the effect of presence on emotion evaluation. Here, we ask how consistently behaviours are rated in terms of valence and arousal under two viewing conditions: (1) the robot is present; (2) the robot is displayed via video.

Of the 72–item behaviour set generated by participants (see *Generating* 4), *CuddleBits Study 2* used a subset of 16: five researchers selected the most representative designs, converging on the top four per quadrant. Under two viewing conditions {live, video}, participants chose three words that best represented the displayed behaviours and rated their confidence in each chosen word, as well as one or more words that least represented the behaviour. Participants rated words ahead of time in terms of arousal and valence. Ratings per participant and per

Figure 5.2: Mean behaviour ratings (+2 for Match; -2 for Not Match) for FlexiBit grouped by the researcher-designed behaviours (horizontal) and the emotion word against which participants rated behaviours (vertical). Researcher-designed behaviours correspond with (1) to (8) in Figure 5.1. RibBit scores were similar and omitted for space.

viewing condition were combined into a single (valence, arousal) point (described below). Through this, we explored the following:

RQ 1. Is there a difference in viewing conditions?
Hypothesis: Participants will rate behaviours similarly regardless of viewing condition.

RQ 2. Are behaviours consistently distinguishable?
Hypothesis: Each behaviour will be distinguishable.

RQ 3. Which behaviour design and waveform features correlate with rated dimensions of arousal and valence?
Conjecture: Features that are characteristic of variability will correlate with va-

lence, while features that are characteristic of speed will correlate with arousal.

5.2.1 CuddleBits Study 2: Methods

Participants: We recruited 14 naïve participants (4 male), aged 22–35. 12 participants were fully proficient in English; the remaining 2 had advanced working knowledge. Out of 14 participants, 13 reported having at least some interaction with pets; 6 rarely interacted with robots, and 8 never interacted with robots. All were compensated $15 per session.

Procedure: Participants were seated, introduced to a fur-covered RibBit, and asked to touch the robot to reduce novelty effects. To calibrate emotion words, participants rated the valence and arousal of 12 words on a 9-point scale (Table 4.2). Participants then viewed the 16 robot behaviours in two counterbalanced viewing condition blocks {live, video}.

In the live condition, participants could physically interact with the CuddleBit while playing each robot behaviour via MacaronBit. Noise-cancelling headphones played pink noise to mask robot noise. In the video condition, participants watched silent videos of the CuddleBit performing the same behaviours (side view, 640x360 px, 30fps). In both conditions, behaviour order was randomized for each participant.

In each viewing condition, participants were asked to choose 3 emotion words that best represented the behaviour from a list of the 12 emotion words they calibrated previously, indicating their confidence level of each word on a 5-point Likert scale. They watched 16 behaviours and answered qualitative follow-up questions. After an optional 5 minute break, this process was repeated, with condition block counterbalanced. Including a semi-structured interview, the session took $\sim$60 minutes.

5.2.2 CuddleBits Study 2: Data Preprocessing

Before each session, participants calibrated the emotion words that they would be using by rating each in terms of arousal and valence. Using the calibrated list of

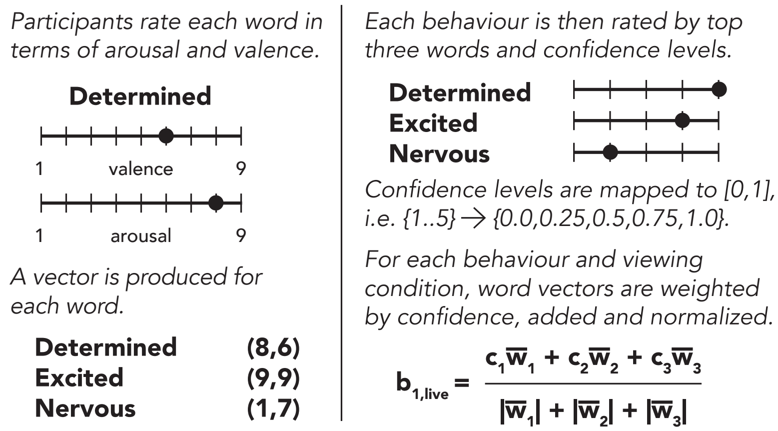


Figure 5.3: For each behaviour and viewing condition, a single vector was calculated by adding the vectors of the top three words that participants chose, weighted by confidence levels. Word vectors were determined at the beginning of the session, when participants rated each word in terms of arousal and valence.

emotion words, we constructed vectors of ($v =$ valence, $a =$ arousal) for each word, where $1 < v, a < 9$. For each behaviour and viewing condition, the best three words were weighted by their confidence values, added and normalized. This produced a single vector of (v, a) for each behaviour and viewing condition (Figure 5.3).

5.2.3 CuddleBits Study 2: Data Verification

Before the following analysis, we ran a series of data verifications to ensure consistency in each participant's responses.

Due to the high subjectivity of the kinds of emotions people will associate with different words, the participant-calibrated emotion words were checked for consistency with the expected PANAS quadrants. For all participants, no more than two words disagreed with the PANAS quadrants; as such, we took the participant rated words to be reasonably calibrated.

Similarly, for each behaviour per view condition, the best three rated words were checked both against themselves, and against the selected least representative word(s). Roughly 50 per cent agreed within a reasonable margin of error across either valence or arousal; 30 per cent agreed across both valence and arousal; 20

per cent either did not agree or were inconclusive.

To determine whether our confidence value weighting scheme was valid, we performed both a visual inspection of word distribution and confusion matrices with design labels. With no weighting scheme, data was heavily biased towards (positive valence, high arousal) ratings, which did not agree with our qualitative results or a reasonable reading of our quantitative results. As such, a linear weighting scheme was determined to be the least biased, such that confidence ratings of $\{1,2,3,4,5\}$ were mapped to $\{0.0,0.25,0.5,0.75,1.0\}$.

5.2.4 CuddleBits Study 2: Analysis and Results

We summarize our findings from *CuddleBits: Study 2*. All significant results are reported at $p < .05$ level of significance.

RQ1: Is there a difference in viewing conditions?

Hypothesis: Participants will rate behaviours similarly regardless of viewing condition. – *Not supported.*

Behaviour label $\times$ **Viewing condition**: We found a significant effect for viewing condition ($Pillai = 0.563$, $F(2,415) = 6.87$) and behaviour label (0.563, $F(30,832) = 10.86$). We did not find an interaction effect ($p = .33$). Although there is evidence to suggest that participants do rate behaviours differently, since they also rate viewing conditions differently, we should be careful in using video as a proxy for live robot behaviour display.

Behaviour label quadrant $\times$ **Viewing condition**: We found a significant effect for viewing condition ($Pillai = 0.441$, $F(6,880) = 41.43$) and by collecting designs by quadrant (e.g., *Hostile* and *Upset* are both high-arousal, negative-valence emotions), ($Pillai = 0.030$, $F(2,439) = 8.705$), and the interaction effect.

Duration: Through 2-way ANOVAs, we found significance in duration between viewing conditions wherein participants took longer to rate live ($\mu = 72.49s$, $\sigma = 40.69s$) than via video ($\mu = 64.13s$, $\sigma = 29.28s$) per behaviour, corroborated in that live behaviours ($\mu = 2.36$, $\sigma = 1.51$) were played more times than the corresponding video ($\mu = 1.96$, $\sigma = 1.18$). The more time spent on live behaviours could be due to more information conveyed or more interest as participants interpret the motion and/or haptic expression.

RQ2: Are behaviours consistently distinguishable?

Hypothesis: Each behaviour will be consistently distinguishable. – *Partially supported.*

Behaviour label $\times$ **Participant**: As behaviours ($Pillai = 0.917$, $F(30,448) = 12.649$), participants ratings ($Pillai = 0.671$, $F(26,448) = 8.705$), and the interaction are all significant, we determine that the behaviours are distinguishable by participant.

Through Figure 5.4, we examine rating consistency by behaviour and quadrant. Negative-valence, low-arousal (*Depressed*) behaviours have the largest dispersion in rating for both dimensions, suggesting that they are the most difficult for participants to classify. Low-arousal, positive-valence (*Relaxed*) behaviours are more consistently concentrated towards the relaxed quadrants.

Both high-arousal, negative-valence (*Stressed*) and high-arousal, positive-valence (*Excited*) behaviours are concentrated in the high-arousal half, yet highly dispersed across valence, suggesting valence is difficult to determine for certain high-arousal behaviours.

Overall, behaviours designed for a representative quadrant may not necessarily be interpreted as such. *Determined*, for example, was interpreted as negative-valence with high-arousal, a contrast to the intended positive-valence high-arousal. Finally, live behaviours (red in Figure 5.4) are more dispersed than video behaviours (blue). This illustrates a higher variation in how participants rated live than video behaviours.

RQ3: Which behaviour design and waveform features correlate with rated dimensions of arousal and valence?

Conjecture: Features that are characteristic of variability will correlate with valence, features that are characteristic of speed will correlate with arousal. – *Partially supported*.

Analysis using machine learning techniques was performed as a preliminary step to understand which features might be most relevant. Using the full set of designed behaviours from participants (see *Generating*) and their associated design labels, we trained a Random Forest classifier on statistical features calculated

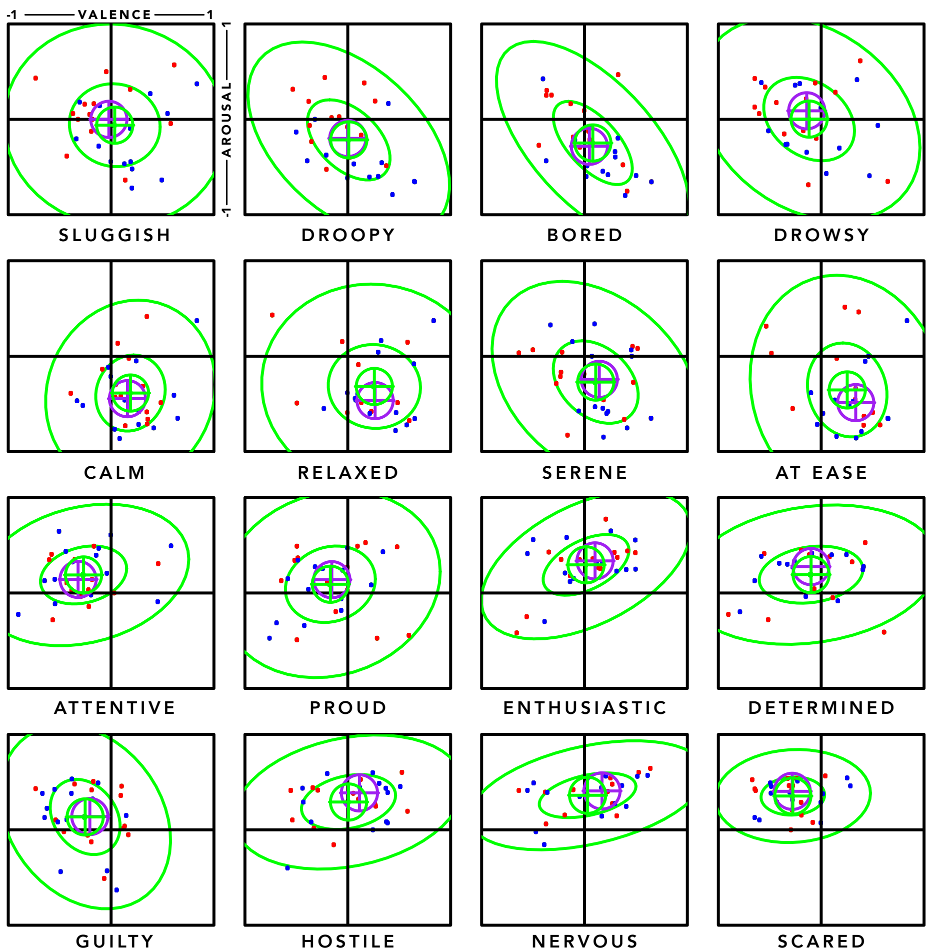


Figure 5.4: Each plot shows a single behaviour's arousal (-1,1) and valence (-1,1) ratings. Live viewing condition is in red, video in blue. Green ellipses show confidence intervals at 5% and 97.5%. Green cross is mean, purple cross is median. Each plot corresponds to a single PANAS word, each row corresponds to an affect grid quadrant. Rows order from the top: *Depressed*, *Relaxed*, *Excited*, *Stressed*.

from design and output waveform attributes. Since each behaviour was output as a waveform, we could decompose the waveform using MacaronBit design parameters, and describe them using keyframe count and standard statistical features (min, max, mean, median, variance, total variance, area under the curve) on keyframe values. The same statistical features were calculated for the output waveform.

Each behaviour label was mapped to the original PANAS quadrant (called here *design quadrant*). When running 20-fold cross-validation classifying on design quadrant, the Random Forest classifier achieved between 66% and 72%, full feature subset for the former, and an optimal subset for the latter (chance=25%). Top performing features were **position**: keyframe count, range, total variance; **random**: max, min.

Note that the selected features are related to waveform complexity. If the *random* parameter was set high, then the waveform would have a high amount of variation. Similarly, if there were a high number of *position* keyframes, the waveform would have a lot of variation.

Feature Selection

A correlation matrix was constructed between arousal and valence for 16 participant-rated behaviours (per viewing condition), and for the 72 participant-and-researcher generated unrated behaviours.

As seen in Figure 5.5, arousal has stronger correlation within the feature vector than valence. Features with strong positive correlation to arousal are those that also correspond with the widest, fastest, and most erratic motions, such as position keyframe count, position range, and random maximum.

Valence has much weaker correlation overall, and particularly low absolute correlation values in the participant-rated analysis. However, within the unrated behaviours, the top correlated features are also indicators of waveform complexity and are negatively correlated with valence, i.e., the more complex a behaviour is, the less it is deemed to be a pleasant behaviour.

Participant Experience

Interviews with participants were audio-recorded, transcribed, and coded for themes and keywords by a single researcher using an affinity diagram and constant comparison. Open-ended written responses from participants for both live and video viewing conditions were analyzed with the same techniques.

In contrast to video, participants emphasized the importance of haptic feedback

	Waveform						Random						Position						
	live_a	live_v	video_a	video_v	unrated_a	unrated_v	live_a	live_v	video_a	video_v	unrated_a	unrated_v	live_a	live_v	video_a	video_v	unrated_a	unrated_v	
	n/a	n/a	n/a	n/a	n/a	n/a	.02	-.01	-.05	.02	-.03	-.19	.39	-.02	.49	-.13	.39	-.15	numkf
	.03	-.06	.09	-.16	-.12	.08	.40	-.10	.46	-.07	.35	-.05	.01	-.16	.02	-.10	-.10	.01	med
	-.03	.07	-.14	.11	-.12	-.21	.43	-.11	.45	-.09	.36	-.17	.13	-.10	.16	-.08	.06	-.25	max
	-.20	.16	-.15	.17	-.10	.17	.28	-.09	.32	-.06	.28	-.18	-.34	.04	-.35	.09	-.24	.18	min
	.26	-.13	.26	-.18	-.01	-.22	.22	-.02	.21	-.02	.23	-.10	-.01	.03	-.04	.12	.05	-.08	var
	-.11	.00	-.16	-.04	.18	-.16	.13	-.08	.05	-.04	.20	-.13	.37	.00	.44	-.05	.49	-.08	totvar
	.15	-.10	.07	-.10	.06	-.24	.25	-.04	.24	-.06	.30	-.12	.34	-.07	.35	-.11	.24	-.31	range
	-.05	.00	-.02	-.07	-.15	.07	.40	-.11	.43	-.08	.36	-.12	-.22	-.08	-.25	.04	-.19	.10	mean
	-.05	.00	-.02	-.07	-.15	.07	.29	-.06	.23	-.04	.23	-.04	.39	-.04	.49	-.15	.36	-.15	auc

+0.5 correlation -0.5

Figure 5.5: Correlation results from behaviours that were designed for an emotion label but unrated by participants (marked *unrated* above) were calculated on all 72 designs from *CuddleBits: Participant Generated Behaviours* (see *Generating*); correlation results from participant-ratings were calculated on the 16 behaviours from *CuddleBits: Study 2* (marked by viewing condition). A strong positive correlation is shown between the position total variance for all arousal columns ($unrated_a$, $video_a$, $live_a$) – the higher the total variance, the higher the arousal.

(7/14), the ability to view the robot from multiple angles (4/14), the increased engagement and accessibility that resulted from the live interaction (5/14); 9/14 touched the robot while playing behaviours. Of these, three reported wanting to touch the robot when they were having difficulty interpreting the behaviour.

> "*Feeling the movements rather than just watching them helped me get a better sense of an interpretation of what the emotion was.*" –P05 (corroborated by P09, P11)

Participants reported that some of the given emotion words were ambiguous (particularly *Depressed*, *Attentive*, *Excited*, and *Stressed*), due to a lack of context and visual cues.

> "*There were some emotions where it was pretty ambiguous. [There are] different connotations depending on exactly what the context is.*" –P03 (P02, P10)

Several participants (4/14) interpreted a combination of emotions while observing the robot behaviours, with emotions happening either simultaneously or sequentially.

> "*I think [the emotions] are happening in order. So sometimes [it's] excited, and then after the stimulus is gone, it becomes bored again.*"
> –P06 (P04, P07, P09)

Participants' description of their process for labelling robot behaviours included relying on their experience with animal behaviours (4/14), their experience with human emotions (2/14), and interpretation of the "heartbeat" or "breathing" of the robot's movement (7/14). 5/14 participants mentioned that it was difficult to interpret emotions from the robot behaviours.

5.2.5 CuddleBits Study 2: Takeaways

Viewing conditions: Experiencing the robot live seems to have a different effect on behaviour interpretation than via video. Since the ratings in viewing conditions were significantly different, it is inadvisable to use video as a proxy for robot behaviours. This difference is likely due to the ability to experience the robot haptically and visually from multiple angles. Futher, the laboratory context may have diminished the emotional interpretation in both cases, as the context of the behaviours were intentionally unspecified and therefore ambiguous.

Distinguishability and consistency: Behaviours were distinguishable; some robot behaviours were consistently rated within the same affective quadrants, some across a single dimension (usually arousal), and some not at all. In line with an intuitive understanding of emotional behaviours, this suggests that it is possible to create behaviours that are seen as meaningfully different (especially in terms of arousal), but that their interpretation is subject to some variance. If we take the distribution of behaviour ratings at face value, the ambiguity of interpretation may not be measurement noise, but a true representation of the behaviour's emotional space. That is to say, the behaviour labeled as *Sluggish* in Figure 5.4 may be well-defined as a probability space with a nearly-neutral mean.

Waveform features: The waveform features that define arousal seem clear: by increasing the amplitude and frequency of a behaviour, it should be interpreted as higher arousal. Valence is less clear, since low correlation values make a definitive claim difficult. However, we conjecture that the waveform features that create a more complex and varied behaviour (i.e., randomness, number of keyframes needed to make a behaviour) create a lower-valence behaviour. More work is needed to look into valence behaviours (see *Complexity* below).

It is possible to create distinguishable emotional behaviours with the CuddleBits. Future work should (1) seek to establish the factors that produce consistent valence interpretations; and (2) establish how those behaviours change within the context of an interactive 'scene'.

5.3 Voodle: Co-design Study

To understand how Voodle would work in a behaviour design process, we performed a co-design study: performer-user input guided iteration on factors underlying Voodle's expressive capacity.

Because using iconic input to generate affective robot motion is an unexplored domain, we focused on rich qualitative data. Methods borrowed from grounded theory [86] allowed us to shed light on key phenomena surrounding this interaction style, and to define the problem space through key thematic events as a basis for further quantitative study.

5.3.1 Voodle: Methods

Ideal Voodle users are performance-inclined designers. We recruited three expert performers to help us improve and understand Voodle.

Over a six week period, each performer met us individually for three one-hour-long sessions, for a total of nine sessions conducted.

After completing Session 1 with all three participants, we iterated on the system for Session 2; we repeated this process between Sessions 2 and 3.

In each session, participants were guided through a series of emotion tasks, followed by an in-depth interview. Each emotion task was treated as a voice-acting *scene*, where the participant played the role of *actor*, and two researchers played the roles of *director* (here, an assistant as for comparative Study 1) and *observer*. As before, the director/assistant offered technical support and suggestions as needed, but did not actively design behaviours. An observer took notes.

In each task, participants used Voodle to act out transitions between opposing PANAS emotional states, e.g., *distressed* $\rightarrow$ *relaxed*, for (high-arousal, high-valence) $\rightarrow$ (low-arousal, low-valence). The full set of emotion tasks (a) crossed the diagonals of the affect grid; and (b) crossed each axis: *Distressed - Relaxed*, *Depressed - Excited*, *Relaxed - Depressed*, *Excited - Distressed*, *Relaxed - Excited*, *Depressed - Distressed*.

Participants performed as many as they could in the time allotted per session. Each session lasted an hour: 30 minutes dedicated to the main emotion task, 20 minutes for an interview, and 10 minutes for setup and debriefing.

An in-depth interview was framed with three think-aloud tasks, to motivate discussion and draw out user thoughts on the experience of voodling. Participants were asked to (1) rate and discuss Likert-style questions of 5 characteristics: perceived *alignment*, *fidelity* and *quality* of designed behaviours, and perceived degree of *precision* and *nuance*; (2) sketch out a region on an affect grid to represent the expressive range of the robot (Figure 5.6). (3) pile-sort [9] pictures of objects, including pets, the CuddleBit, and tools, to expose how they defined terms like 'social agent,' and how the Bit fit within that spectrum.

5.3.2 Voodle: Participants

Participants were professional artists with performance experience, recruited through the researchers' professional networks.

P1 is a visual artist focused on performance and digital art. He was born in Mexico and lived in Brazil for 4 years and Canada for 7 years. P1 is a native speaker of Spanish and English, with working knowledge of Portuguese and French.

P2 is an audio recording engineer, undergraduate student (economics and statistics), and musician: he provides vocals in a band, and plays bass and piano. P2 is a native English speaker born in Canada; he is learning German and Spanish.

P3 is an illustrator, vocalist, and freelance voice-over artist. She has a degree in interactive art and technology, and has taken classes on physical prototyping and design. She is a native speaker of Mandarin and English, with working knowledge of Japanese. P3 was born in Taiwan and immigrated to Canada when she was 8 years old.

5.3.3 Voodle: Analysis

We conducted thematic analysis [73] informed by grounded theory methods [24] on observations, video, and interview data. We found four themes (Table 5.1): participants developed a *personal language*, voodling requires a *narrative frame* and brings users into *alignment* with the robot, and parametric *controls* complement the voice for input. Each session helped to further develop and enrich each theme, adding to the overall story. We refer to each theme by an abbreviation and session number: "PL1" is Personal Language, Session 1.

Table 5.1: Summary of *Voodle* Co-Design Themes. We refer to each theme by abbreviation and session number (e.g., "PL1")

Theme	Definition	Session 1	Session 2	Session 3
Personal Language (PL)	The individualized words and utterances a participant developed with the robot.	Participants took a varying amount of time to "get" Voodle; each vocalized in different ways, arriving at a local maximum.	Participants build upon their constructed language, starting from their Session 1 language, but exploring more ideas.	Robots influenced choice of voice or MIDI input, but not vocalization language.
Narrative Frame (NF)	The story the user is telling themselves about who or what the robot is.	Participants needed to situate the robot by constructing a character to effectively interact with the robot by utilizing metaphors, concepts, and feelings that do not need to be explicitly described in words.	Participants used narrative frame in different ways. Fur did not affect their ability to construct a narrative frame.	Robot form factor, orientation adjusted the stories that participants told.
Immersion (I)	The extent to which a participant could suspend their disbelief.	Participants adjusted their language depending on the robot's behaviour. By *conversing* with the robot, they found the behaviour was more believable than the observers did.	Experience helped people be more in-tune ("aligned") with the robot; as did voodling in comparison to using direct MIDI controls.	Too much or too little control reduces emotional connection; physically actuated displays connect more with users.
Controls (C)	How the control of the system influenced how participant saw the interaction.	Laptop controls were difficult to use. A low pass smoothing algorithm was not effective. Randomness contributed to life like behaviour.	Physical MIDI controls were easy to use when voodling, but lacked feedback. The robot needed an adjustable "zero" to maintain lifelike behaviour without input.	Suggestions include: steady-state sine wave breathing and setting 0 position as 50% of max servo

Voodle: Session 1

We introduced naïve participants to the initial Voodle prototype and allowed them to explore its capabilities and limitations by completing as many of the emotion tasks as time permitted ($\sim$30 mins). We closed with a semi-structured interview; participant feedback informed the next iteration of Voodle.

Theme PL1: From "Eureka" to local maximum: Participants were initially instructed to use iconic vocalizations with an example such a 'wubba-wubba'. Despite this, all participants chose to use symbolic speech early in Session 1.

For example, when asked to perform the emotion task *relaxed* $\rightarrow$ *depressed*, P2 started by saying *"I'm having a nice relaxing day"*, with little visible success in getting the Bit to do what he wanted. Each participant transitioned into understanding how to use Voodle at different times. P3 quickly understood that symbolic speech wouldn't afford her sufficient expressivity, and transitioned to iconic input, while P2 kept reverting to symbolic speech as an expressive crutch.

It took P2 until the fifth emotion task (of seven) until he had a breakthrough: *"I kinda made it behave how I imagined my dog would behave"*. Using that metaphor, subsequent vocalizations attained better control. Unlike the other two participants, P1 switched to iconic vocalizations gradually.

Each participant eventually converged on his or her own idiosyncratic collection of sounds that they felt was most effective. This differs from what might be a globally optimal set of sounds to use: participants stayed in some local maximum. For example, P1 started using *"tss"* sounds and breathes into the microphone; while initially successful for percussive movements, they later proved limiting. P2 used nasal sounds peppered with breathiness (*"hmmm"*). P3 eventually focused on manipulating pitch with vowels. (*"ooOOOO"*), as well as employing nasals like P2 (*"mmm"*), and some ingressive (breathing in) vocalizations. (*"gasp!"*).

Theme NF1: Developing a story: Once the participant finds the robot's 'story', emotional design tasks get easier. For example, P2's shift came with his story of the robot being a dog. P2 refused to explicitly tell a story: *"it wasn't much of a concrete story"*. P1 said he created less a full story, *"more a grand view of feeling some emotions and from there on you could build a story, we were getting more the* *traces of a story through the emotions"*. This narrative potential was enabled by the conceptual metaphor of *Voodle as a dog* [47].

Theme I1: Mirroring the robot suspends disbelief:

Participants formed a feedback loop with the robot: their vocalizations influenced the robot's behaviour, which in turn encouraged participants to change their vocalizations. P2's dog-like *"hmmm"* vocalizations caused the robot to jitter, surprising P2 and prompting a switch to *"ooo ooo ooo"* sounds.

When actively interacting with the robot, participants reported stronger emotional responses than the experimenters observed in the robot; as *actors* in the scene, participants were more connected than the *director* and *observer*. This could be due to their close alignment with the robot while acting – an experimenter might see a twitch as a quirk of the system, but the participant might see it as evocative of emotional effort: *"I did an 'aaa' and at the end of the syllable it did a flutter...it was just really nice, there were just things that I didn't expect that expressed my emotion better than I thought it would"* (P3).

Theme C1: Screen distracted, algorithm was unresponsive: Using a laptop to control algorithm parameters distracted participants from looking at the robot.

All participants had some trouble modifying Voodle parameters; the director needed to take over parameter control (with the participant's direction) as they vocalized. Parameter manipulation was especially difficult in emotional transition tasks where multiple parameters needed to be adjusted over time.

In addition, participants reported that the smoothing algorithm, a simple low-pass filter, was unresponsive: *"feels like there's a compressor [audio filter restricting signal range]...limiting the the amount of movement"* (P2).

Changes for Session 2 – We implemented four changes for Session 2: replaced the web interface with a physical MIDI keyboard for parameter control; replaced the low-pass filter with a PD (proportion/damping) controller to improve responsiveness, with parameters named "speed" (P) and "springiness" (D); introduced a new "randomness" parameter to simulate the noise from the removed low-pass filter; and added a new mode to aid in making comparisons, "Wheel": users could press a button on the MIDI keyboard to disable voice input and directly control the position of the robot using wheel control.

Voodle: Session 2 Format and Results

In the second session we juxtaposed voodling against a manual MIDI-wheel controller, based on a participant's suggestion. Participants first did as many emotion tasks as possible in ∼15mins, using voice for robot position control; then repeated these in Wheel mode. Included in this session were observations of how a user's relationship with the Bit matured as they became more familiar with both the robot and Voodle.

PL 2: Participants learn, differ in skill: Unprompted, participants began with same language they used in Session 1, then developed their language with experimentation. P3 continued to use primarily pitch control, as she did in the first session. P1 continued his *"tss"* sounds and blowing directly into the microphone, essentially a binary rate control: the robot was either expanding quickly or contracting quickly.

After some experimentation, P1 incorporated more pitch control, which afforded better control. P2 and P3 indicated increased expressivity on their affect grids in Session 2 (Figure 5.6), suggesting improvement of either ability or system.

The participants began to diverge in their ability to create nuanced behaviours, suggesting talent or training influenced their capabilities.

P1, with his breathing sounds, simply didn't succeed in controlling the robot. P3 seemed to understand how to work with Voodle, creating subtle and expressive designs; she preferred vocal input, but also was adept with Wheel. P2 was between the other two, making extensive use of Wheel control, and playing it like a piano.

NF 2: Agency from motion:

Randomness and lack of precise control imbued the robot with agency. P1 claimed that, on the whole, randomness made the Bit feel more alive because it implies self-agency. When turning up the randomness, P3 exclaimed, *"oh hey hi, I woke it up"*. She explained: *"The randomness meter...was always the first thing I moved I think...because it added another layer of emotion to it."* This lack of control connected to the sense of life within the robot: *"[the Bit] was modeled to look like a living creature and that makes me feel like it should probably not completely obey what I want it to do. There should be something unexpected"*

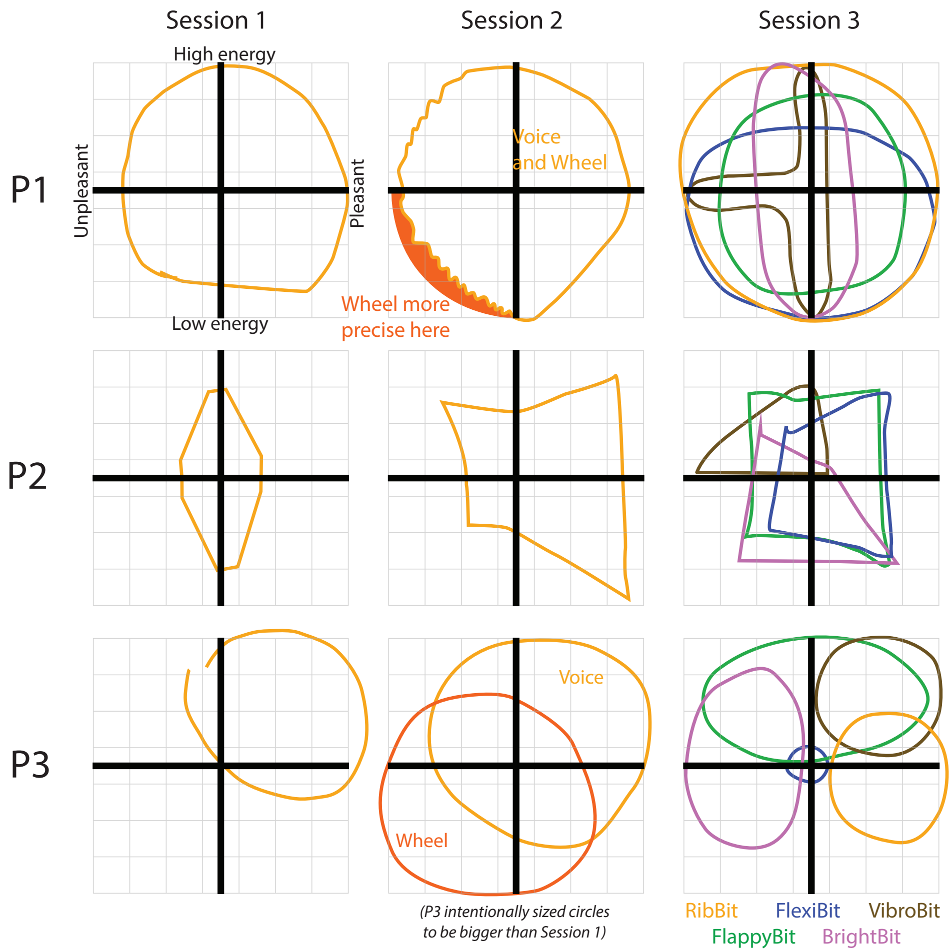


Figure 5.6: Reported affect grids by participant and session. After being instructed about dimensions of arousal and valence, participants drew the robot's expressive range directly on affect grids. Participants indicated increased expressivity from sessions 1 to 2, differences between voice and Wheel control, and that each robot had a different range.

(P3).

Continuous motion can contribute to agency. All participants felt the robot should not be motionless in its 'off' state; it needed a default, like breathing. P2 further suggested that the robot's 'zero' point be the middle of its range, to accommodate both contraction and expansion metaphors.

I2: Voice converses, MIDI instructs: Participants were more aligned with the robot when vocalizing.

For example, P3 expressed that manual wheel control allowed her to instruct the robot, whereas voice control allowed her to converse with the robot: *"Voice feels like it's more conversing than by wheel, I think it's because by wheel I have a better idea of what's going to happen...which makes me experiment a little less"* (P3). Non-voice MIDI control gave a stronger sense of controlling the robot, diminishing agency: *"[The wheel] felt more like playing an instrument"* (P2). P2, the audio engineer, preferred using the MIDI wheel, while P3 preferred voice. Both P1 and P3 indicated that the Wheel had more expressive capabilities with low-arousal, negative emotions (Figure 5.6).

C2: Visual parameter state: MIDI parameter control allowed participants to focus attention on the robot.

All participants continuously modified the Voodle parameters with the MIDI controller, compared to minimal modification with Session 1's HTML controller. P2 suggested that sliders may be more effective than knobs, as they provide immediate visual feedback for range and current value. P3 also requested more visual feedback for parameter status, e.g., bar graphs.

Changes for Session 3 – We displayed parameter status on the laptop screen, and added 4 new robot forms to explore how form and actuation modality influence voodling (Figure 4.3).

Voodle: Session 3 Format and Results

The final session explored the effect of form factor on control style. Each participant ran through a subset of our emotion tasks with each of the new robots (Figure 4.3), given the option to use either voice or wheel control. They were also allotted free time to play with the new robot forms. We administered a closing questionnaire to capture their overall experience of the final version of Voodle.

PL3: Consistent language across robots : Despite wide variation in each robot's expressive capability (Figure 5.6), participants continued to use their developed languages across robots. Examples include *"tssss"*, *"ooo"*, *"aaa"* (P1), *"mmmm"*

(P2), and *"oooh"*, *"ahhh"* (P3). While language remained consistent across robots, preferred control mechanism did not.

P2 preferred vocal input only for FlappyBit as he engaged emotionally with it: he saw the flapper as a head. However, P2 used wheel control for the remaining Bit forms. P1 always started vocalizing as an experimentation technique with new Bit forms and then consistently moved to wheel input for fine-grained control. P3 preferred voice for most robots, although she did indicate the RibBit responded more consistently to wheel input (unlike the other robots).

NF3: Shape, orientation create lasting stories: Robots did not just have varying expressive capability; they also inspired different stories. Participants reacted differently to each. For example, P3 saw VibroBit as a multi-dimensional, highly-controllable, lovable pet; P1 and P2 saw it as a unidimensional, completely uncontrollable, unlovable object. Different robot features changed the narrative context. While P2 thought FlappyBit's flapper was a head, giving it expressivity, P3 thought the flapper was a cat's tail. When FlappyBit was flipped over such that its flapper curled downwards, both P2 and P3 felt that it became only capable of expressing low-valence emotions. However, form factor did not not completely change the story: in all sessions, P2 felt the robot was a dog, no matter which robot he was interacting with.

I3: Sweet spot of control; motion matters: P1 reported high control over BrightBit and low control over VibroBit, but rated both with a smaller expressive range than the other robots (Figure 5.6). This suggests a "sweet spot" of control when connecting emotionally with the robot: some control over behavior is good, but not too much.

P1 felt more connected to FlappyBit or FlexiBit. That said, all participants expressed a lack of emotional connection with BrightBit. P3 thought that the lack of movement was the cause, while P2 did not feel like he conversed with BrightBit: *"I kept visualizing it talking to me instead of me talking to it"* (P2).

Changes for Robot Iterations – Session 3 resulted in several implications for future iterations on each robot: VibroBit had a limited expressive range; FlappyBit's flapper looked like a head, which was easy to connect with, but metaphors would vary depending on orientation; FlexiBit had an ambiguous shape; BrightBit seemed unemotional.

Voodle: Likert and Pile-Sort Results

The Likert scale and pile sort tasks were primarily used as an elicitation device to stimulate discussion. Participant responses were consistent with other observations; we highlight a few examples.

The questionnaire measured *quality* of Bit movements match to participant's vocalizations/manual control; *precision*, *nuance* and *fidelity* of voice control; and *alignment* of Bit behaviour to the emotions participants felt as they performed.

Emotional connection with the RibBit increased by session. RibBit and CurlyBit performed much better than other CuddleBit forms on all metrics. Wheel and voice control offered similar degrees of *quality* on average. P1 and P2 reported that they felt more in control with the wheel, though P3 said that it made the Bit appear as less of a creature.

Participant perceptions of the CuddleBit as a social agent changed through repeated sessions, albeit in different ways. In the pile sort, P2 first placed RibBit between cat and robot, but post-Session 2, moved to between human and cat. In contrast, P1 first sorted the RibBit between a category containing anthropomorphic elements and home companionship possessions, but later agreed it could fit in all of his categories (except one for food) if it was wearing fur.

5.3.4 Voodle: Discussion and Takeaways

Our initial goal was to create a dedicated design tool for affective robots. We observed something intangible and exciting about live vocal interaction. We derived a more nuanced understanding of Voodle use, in that it seems to exist somewhere between robot puppetry and a conversation with a social agent.

In the following, we discuss insights into interaction and believability, and how Voodle can function as an interactive behaviour *design tool* within a performance context. We conclude with future directions, including insight into how Voodle might be embedded as a component of a larger behaviour control system.

5.3.5 Voodle: Insights into Believability and Interactivity

Through the co-design study, we found that believability was mediated by participants conception of robot *narrative context*, and their *level of control* and *personal ways of using* it.

Creating a context:

Behaviour designs and alignment improved dramatically once participants found a metaphor or story. Context was determined by confluence of form factor, robot ability and participant-robot relationship. For example, P1 could neither decide what VibroBit represented nor control it well, hence saw it as a failure; while P3 thought that it was cute and felt skillful when interacting with it.

Balancing control with a "spark of life": Voodling created lifelike behaviours with a simple algorithm: deliberate randomness and noise produced a user-reactive system that still seemed to act of its own accord.

Varying randomness and user control made Voodle more like a conversation, or like a design tool.

Control increased *alignment*, like people sharing mannerisms in a conversation [33]; but with too much or little, the system becomes mundane or frustrating, the magic gone. Voodle was a more emotionally immersive design experience than traditional editors.

Personalization:

Users developed unique ways to use Voodle. Algorithm parameters could be varied to facilitate a metaphor, output device, or simply preference. Users modulated their vocal performance with these parameter settings much as guitarists use pedals to adjust tone, before or as they play. Importantly, we observed that users tended to use similar "personal language" with varied robots, suggesting an individual stability across context.

Voodle: Vision for Behaviour Design Process

It is likely that producing affective robots will soon be like producing an animated film or video game. Indeed, steps towards this have begun (e.g., Cozmo [3, 36]). Here, it seemed that enabling artists and performers to directly interact with robots during design did facilitate the believability of the resulting behaviors, in that the designers who became aligned with their robot model seemed to be more satisfied with their behaviors than more attached observers.

Behavior design team: As reflected in the structure of the co-design study, a behaviour design session may involve a scripted scenario, a director, a designer, an actor, and the robot itself. Working together to bring out the best performance on the robot, an actor and director would read through a script as the designer takes notes on how to modify the robot's body. Through an iterative design process [16, 40], both behaviours and robot form factors could be refined together (Theme NF3).

The actor could also leverage Voodle's support to improve *alignment* with the robot. Like a puppet, the actor would be simultaneously controlling and acting with the robot. Although the interactive space in which the actor works will likely have to be multimodal (i.e., , including a physical controller such as the MIDI keyboard), alignment through voice enables a deeper emotional connection with the robot itself (Theme I2).

Physically adjustable parameters: Voodle took a different approach from previous non-speech interfaces (e.g., the Vocal Joystick [10, 37]), which had a defined, learned control space. As we discovered in our pilot, voodling relied on a narrative context: a metaphor for how vocalizations should produce motion. This could change from moment to moment: amplitude might be associated with the robot expanding, but if the robot was conceptualized as "flexing", amplitude corresponded to downwards movement. When adding parameters, we found physically manipulable controls were easier to control when voodling, but they require visual indicators of their range and status.

One could imagine a kind of *recording engineer* in a behaviour design session who adapts motion control parameters on the fly (Theme C2).

How to Extend

This work produced initial requirements for a Voodle system, which is open-source and online. It also produced implications for future iconic speech interfaces.

Extending the sound-symbolic lexicon:

Here we considered proportionally-mixed pitch and amplitude.

Our pilots (Table 4.1) have already revealed other promising vocal features, such as -continuants (*“dum dum”*), +stridents (*“shh”* or *“ch”*), and distinguishing voiced consonants (*“b”* is voiced, *“p”* is not). A detailed phonetic analysis will highlight additional features and inform ways to adjust parameters automatically for specific vocal features.

Some parameter ranges should be individually calibrated, e.g., pitch.

While we identified examples of our performers’ languages, many more iconic mappings (features to robot position) are possible. These features could further be dynamically mapped to multiple degrees of freedom.

Design techniques: While Voodle was built as a design tool, in context, we found it was rarely used alone. Instead, Voodle could be part of an animation suite, letting users easily sketch naturalistic motion without a motion capture system. Input could be imported into an editing tool for refinement. This might be especially viable in mobile contexts, to sketch an animation on the go, e.g., in a chat program.

Iconic vocalizations have also been used to describe tactile sensations [77, 91], so Voodle may also be useful for end-user design of tactile feedback, to augment communication apps – a haptic version of SnapChat or Skype, with voice for haptic expression. We expect such uses will need to recognize additional linguistic features (like *“sss”* vs *“rrr”*); and Voodle must be more accessible to end-users who are not performers.

Vision for “Embedded Voodle”: Voodle has the potential to add life-like responsiveness to deployed interactive systems. Adding randomness to an ambient display increases perceived agency [8], but voodling could increase a sense that it is *attending* to the user, especially with directed speech (I2).

As a reactive system, voodling could be added to conventionally planned motion of virtual agents or robots, from a robot pet that reacts to ambient speech, to body language of a assistive robot arm (Figure 5.7). When a user explicitly tells the robot arm to *“come here”*, she might modulate its movement with a soft *“whoa”* (*slow down*) or urgent *“WHOA”* (*stop*).

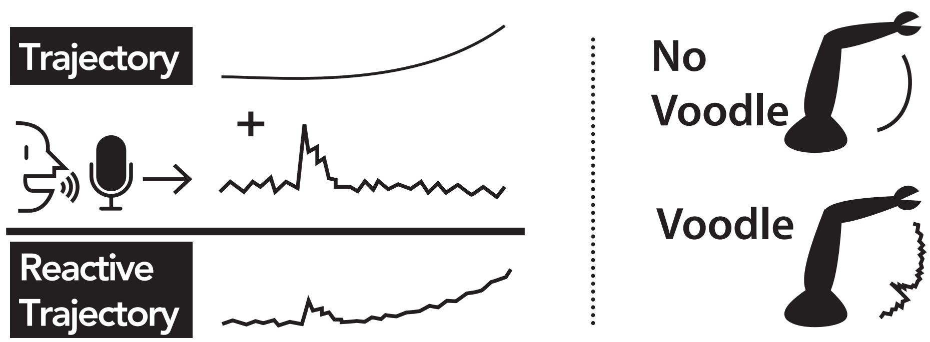


Figure 5.7: Vision for "Embedded Voodle": Voodle could be a natural low-cost method to emotionally color a motion path in more complex robots.

5.4 Complexity: Valence and Narrative Frame

During *CuddleBits Study 2*, we found that certain measures of complexity seemed to be negatively correlated with valence. During all of *Voodle*, *CuddleBits*, and other investigations we found that the 'story' people told about the robot—e.g. *the robot is like my cat* or *the robot is a combination of a dog and a squirrel*—heavily influenced their perception of the valence of the robots behaviours.

This study proposes to generate robot behaviours that vary in complexity and test the participants perception of displayed valence, with consideration for the story that people tell about each behaviour.

Note: in previous work, we have noticed that arousal and valence are not perfectly orthogonal. This means that arousal and valence are possibly dependent, such that an increase in arousal may imply an increase in valence. We have also seen that it is relatively easy to control the perceived arousal of the robot by increasing, for example, the range of the robots breathing. As such, we are targeting valenced behaviours only in this study, using a simple five-point scale from negatively valenced, to positively valenced.

RQ 1. How does valence correlate with complexity? *Hypothesis: valence will be negatively correlated with complexity, i.e., the more complex a behaviour is, the more negatively valenced it will be perceived.*

RQ 2. How do participants rationalize the change in the robots behaviours (i.e., how does the 'story' change the perception of the robot)?

5.4.1 Complexity: Methods

Participants: We recruited 10 nave participants, all compensated \$15 per session.

Procedure:: Particpants were seated, introduced to a fur-covered FlexiBit, and asked to touch the robot to reduce novelty effects. They were then shown a sample behaviour and walked through the rating task, and the story task. They were then asked to complete the rating tasks while being directed by a pre-recorded voice while an experimenter took notes. Each task took roughly 2 minutes: the robot would breathe neutrally, a test behaviour would be displayed, the robot would return to neutral breathing, a pre-recorded voice would ask them to rate the behaviour (a replay option was available), and a pre-recorded voice would ask them to describe the robot's behaviour out loud in a few short sentences. There were twenty trials in total; the first two trials were always the same and were thrown out to reduce novelty effects, the next 18 behaviours were presented randomly. A short follow-up interview ended the session; each session took roughly 45 minutes.

Fifty-four behaviours were designed by hand, then scored and ranked for complexity using three complexity measures: variance, peak count of a spectrograph as generated by a fast Fourier transform (FFT), and MSE. **Variance** was calculated by the standard statistical method, i.e., $E[(X-\mu)^2]$, and was included to compare against results from *CuddleBits*. **FFT peak count** was calculated by counting all maxima on a spectrogram within 2.6 standard deviations of the global maximum, and gives an estimate of the distribution of power across the possible frequencies of the signal. **MSE** was calculated as outlined in *Related Work* (2), and in Costa 2008[26], then behaviours ranked by slope and shape of the produced MSE graph. Six behaviours were chosen from each complexity measure to represent an even spread across rankings, i.e., one from each of the $0-16, 17-32, 33-48, 49-64, 65-80, 81-100$ percentiles for each of $\{$*Var, FFT, MSE*$\}$. This ensured a wide range of possible behaviours were shown to participants.

5.4.2 Complexity: Results

Here, we outline qualitative and quantitative results. Quantitative results outline the consistency of behaviour ratings in terms of valence; qualitative results discuss the impact of narrative frame on behaviour interpretation.

Complexity: Quantitative Consistency

Inter-rater reliability: This measures the extent to which different raters (participants) rated all behaviours similarly. For example, if all participants rated each behaviour the same, there would be perfect agreement. If half the participants rated all behaviours one way, and half the other way, there would be systematic disagreement. Since we are using ordinal scales, a reliability measure that accounts for the order of the scale items was used. For example, we would want to assign a closer agreement between a behaviour rated at both "-2" and "-1" than if it were rated at both "-2" and "+2". Krippendorf's alpha was therefore used to determine the inter-rater reliability, producing $\alpha = 0.07$. This is typically interpreted as *slight agreement*, where $\alpha < 0$ is *systematic disagreement*, $\alpha = 1$ is perfect agreement, and $\alpha = 0$ is no agreement. However, the observed agreement (p_a) is 0.80, and the agreement due to chance is 0.79 (p_e), so α may be suppressed in this case ($\alpha = (p_a - p_e)/(1 - p_e)$).

Per-behaviour consistency: This measures the extent to which each behaviour was rated similarly by all raters (participants). For example, if every participant rated a behaviour differently, we would say it was inconsistent. Weighting ratings for an ordinal scale [45], observed agreement per behaviour (p_i) ranged from 0.64 to 0.92, with 4 behaviours <0.70, 6 behaviours between 0.70 and 0.85, and 8 behaviours above 0.85. As intuition, the variance of ratings and $p_{i:n}$ are highly correlated; the lower the variance in rating per behaviour, the higher the p_i.

Correlation:

Our goal was to explore how measures of complexity correlate with valence ratings of breathing behaviours (i.e., to see whether more complexity produced lower valence).. Since our per-behaviour ratings were reasonably consistent, we consider mean, median, and mode of the valence ratings per behaviour and cor-

Table 5.2: The correlation between summary statistics of valence ratings (columns) and complexity measures (rows). For example, the top left cell is the correlation of the mean valence rating per behaviour and the variance of the behaviour signal.

Valence/ Complexity	**mean**	**median**	**mode**
Variance	-0.03	-0.15	-0.08
FFT peak count	-0.01	-0.12	-0.14
MSE rank	-0.45	-0.47	-0.37
MSE mean	-0.30	-0.08	-0.30
MSE AUC	-0.30	-0.27	-0.32
MSE slope	-0.30	-0.31	-0.43

relate with our complexity measures. For the signal *variance* and *peak count*, we simply calculated the correlation between each summary statistic of valence ratings and the complexity measure of each signal; because MSE outputs a series of values per signal (i.e., one per time scale over which MSE performs coarse-graining), we correlate each summary statistic of valence ratings with the mean MSE rating, area under the curve of MSE ratings, and slope of the regression line of MSE ratings. Results are reported in Table 5.2.

Complexity: Narrative framing

At the end of each trial, participants were asked to describe the robot's behaviour in a few short sentences. Individual interpretation of the robot's character, motivations, and subjective state were highly varied, even within the same behaviour. Here, we present a series of themes developed by (1) comparing and contrasting responses per behaviour; (2) comparing and contrasting responses across behaviours; (3) drawing insights from individual responses.

The qualitative results paint a much more complex picture than the quantitative results. Participant understanding of the robot and its behaviours were highly individualized. Although there were many behaviours for which at least some participants converged on the interpretation of the behaviour, the majority of the responses revealed that the final reported valence measurement was fragile[1] to a

[1] Here, I use fragile to mean that the final outcome could vary widely across the range of possible

complex emotional reasoning scheme. Part of that emotional reasoning scheme I refer to as the narrative frame, i.e., the set of assumptions about the robot's character, state of mind, backstory, and relationship to the environment through which a participant will interpret a behaviour. Here I outline three themes that attempt to characterize the complexity of the reasoning behind the participants' emotional interpretations:

Participants ascribe complex mental states to the robot:

The robot was understood to have thoughts and feelings that extend far beyond it feeling "good" or "bad". Generally, participants gave a nuanced explanation for each behaviour, where they attributed both human- and animal-like motivations and emotions to the robot. For example,

> "It feels like it was trying really hard to do something but it doesn't seem to be able to do it." (P7, B11)

> "It feels calm, but also like he wants to start moving more, like wants to play or something." (P5, B10)

Both quotes explain the current behaviour in terms of the robot *wanting something* that it is not currently able to do, which suggests that the robot not only has a current state of mind, but the ability to reason about the future. Similarly, participants describe the robot having hidden feelings that oppose the current action:

> "I feel like the robot is angry and wants to punch something like that, but it's not like he's lost his mind, he's still under control." (P9, B4)

> "...it kind of felt like it was sort of aggressive response, like someone trying to hold in their anger, or whatever it might be. But perhaps aggressive is not the best word because it didn't seem like it was some sort of...it seemed more like frustration, that's a better word for it, frustration." (P3, B5)

---

responses depending on a few small changes to the emotional reasoning that was used. In contrast to uncertainty, fragility means that the participant could be very certain in what the final valence rating should be, but that conviction could change very easily given a small change in one of their assumptions about the robot.

If the robot is *frustrated* and trying to *hold it in*, it would have to have a sense of how its actions impact the people with whom it is interacting, i.e., it must have some kind of social understanding. This implies that participants not only act as if the robot has an internal mental state, but that the robot has its own internal models of *their* emotional states. This is not naively done; critically, participants are fully aware of the robot's status as a robot but act as if it has these complex mental states regardless.

Narrative frame heavily influenced valence ratings:

Assumptions that participants made about the character, motivation, and situation of the robot heavily influenced whether they saw the behaviour as positively or negatively valenced. For example, both of these quotes describe the same behaviour:

> "The shaking so fast like the robot is talking to his friend like so happy, so makes me feel like it's a positive emotion." (P9, B5)

> "Breathing rate's really fast, seems like it's really anxious, or it's running away." (P2, B5)

In both cases, the story of the behaviour (whether it was running away, or talking to a friend) flip the participant's understanding of the behaviour's valence. Interestingly, both frame the behaviour as if the robot was not reacting to the narrative of the interaction within the lab, but some other, unseen narrative.

Even when all participants gave similar valence ratings, the emotional content of the rating differed greatly. For example, participants who rated B7 as negative framed the interaction as all of having "trouble breathing" (P2), "sobbing" (P3), "alarmed" (P4), "didn't want to be kept here" (P5), "worried" (P6), "agitated". Positive ratings included "restless, but in a good way" (P8), "working [a job]" (P7), "a cat playing with straw" (P1)[2].

Participants often framed the current behaviours relative to previous behaviours, often with short references to "this time" or "last time". One participant even saw a throughline between multiple behaviours:

[2]P9 rated the behaviour positively, but could not decide what the robot was doing.

> “OK, I believe that the last two trials and this trials is a whole story. Like, the first one should be like the robot gets sick at the beginning phase, and the second one is ’the robot is so sick’, and this one, the robot is dying, but he’s still struggling so I feel like a little breaths in-between but almost like, in the most of time, he just doesn’t breathe at all. So maybe he’s dying. So that’s my guess.” (P9, B2)

This was part of building up the robot’s story over time. One participant made explicit reference to the robot’s back story, i.e., where it came from:

> “It seemed like it was burrowing, burrowing into my arms I guess, but I imagine that that’s an instinctive behaviour it got from the wild.” (P4, B10)

Valence was difficult to determine in incongruous or neutral states:
The majority of responses indicated that it was easy for participants to find a story for the behaviour. However, if they could not determine a story, or if the actions seemed incongruous with the story they had already imagined, then it was difficult for them to rate the valence of the behaviour:

> “It didn’t seem like an animal behaviour, breathing was a bit weird. I don’t know—it doesn’t seem like an actual animal behaviour that I recognize. The breathing was just...yeah, it didn’t seem very real to me at all, so I wasn’t able to place a behaviour for that.” (P10, B14, valence= 0)

> “I don’t know, I couldn’t tell what was going on. It seemed pretty random to me, and I couldn’t tell if it was an emotion.” (P1, B16, valence= 0)

In both of the above cases, other participants were able to produce a strong narrative for the behaviour:

> “The behaviour reminded me of a dog gnawing his teeth in his sleep, so this, like, grinding his teeth, and doing, yeah, just, weird things in his sleep. So I gave it neutral.” (P8, B14, valence= 0)

> "Compared to all of the other behaviours, this one, to me, definitely had more of a truly utter joy and satisfaction to it, I don't know why, I can't exactly put my finger on why I think that, just my natural gut instinct is that, just kind of an air of warmth and happiness and satisfaction just the way that the pulse was." (P3, B16, valence= 2)

Often, even with a strong narrative, a neutral state made it difficult for participants to rate in terms of valence:

> "I feel like the robot is aruging with somebody or something and his emotion gets more and more...like, I don't know how to describe it, but he gets more excited while he's doing this, so I think maybe he's trying to convince somebody of something, but I don't think it's either positive or negative, so I choose neutral." (P1, B11, valance= 0)

For example, this quote illustrates a strong (if not well-articulated) sense of what the robot is doing, but not a strong sense of whether the robot is in a positive or negative emotional state. The participant describes the robot as "arguing" and "excited" but cannot determine a valence rating.

5.4.3 Complexity: Discussion, Takeaways, and Future Work

Complexity measures: Higher correlations between MSE measures and valence ratings relative to variance and FFT peak count suggest that MSE is a promising approach for quantifying valence. Future work should include an MSE analysis on the behaviours from *CuddleBits* combined with other complexity measures such as time irreversibility.

Generating behaviours: A generative approach where MSE is used as a utility function might still be possible, but computationally efficient stochastic processes need to explored to seed behaviour generation. Our attempt at generating via genetic algorithms using MSE as a utility function was short lived due to the length of time generation was taking with purely random recombination, even with hand-generated behaviours as seeds. One could imagine using a stocastic process such as Brownian motion or Perlin noise to generate a complex seed behaviour (or portion thereof), then use known complexity-reducing functions to vary the complexity towards some MSE setpoint. However, further tests are still needed to determine how appropriate MSE measures would be for all low-valence behaviours.

Generating on the fly: An interactive robot needs to generate behaviours on the fly. It may be possible to mix behaviours to mix valence states, e.g., through dynamic time warping [23]. More work is needed to know whether a fully-interactive context would significantly change the interpretation of the behaviours.

Qualitative Analysis

Qualitative results present a complex and nuanced picture of robot behaviour interpretation. Although there were some behaviours that were unambiguously positively or negatively valenced, the diversity of stories the participants told about the robots belie a fragile system of interacting narrative elements that could radically shift the understanding of a behaviour.

A natural urge for a scientist is to argue that we need higher experimental control to properly determine valence rating; however, there are are a number of problems with this premise.

First, the dimensions of control are ambiguous, since results indicate that we would want to control the participant's narrative frame. It is not clear whether we can, would want to, or be able to provide a story for the robot, as the story is produced in conjunction with the participant's own subjective experience. The laboratory setting provides its own set of expectations and narrative possibilities; the ambivalence of some interpretations might have been in response to a rating task that had low ecological validity. However, increasing the ecological validity by, e.g. moving out of a laboratory, would likely mean giving up control in many other dimensions.

Second, getting a clear rating of valence may take developing a relationship between the robot and the participant. This would necessitate creating an interactive environment where the robot's actions could vary directly in response to the participant's actions. Although such a situation would make procuring ratings difficult (when and how would participants rate a behaviour?), we may find that some consistent narrative frame emerges naturally out of continued interaction.

Third, it may be that a single-sample, single-point rating task is too reductive. Asking a participant to choose a single point on a dimensional model that represents an entire behaviour ignores any changes in emotion that the participant may have perceived during the behaviour. For example, imagine interacting with the robot for a five-minute period: there may be a lot of different emotions felt in that time, all with different intensities. Further, a single point on any dimensional model may not well represent an emotion state—it seems reasonable to be both "happy" and "sad" at the same time. For example, an emotion state may be better represented by a probability distribution across a set of dimensions, as that would account for simultaneous conflicting emotions. It may be that an emotion state should not converge to a single point until a decision is forced.

Future work should include tasks that account for these difficulties. Accounting for the complexity of subjective measurements, the tasks should (a) attempt to produce a narrative frame through interaction; (b) have some ecological validity through creating a believable setting; (c) take place over a long enough period of time such that a relationship can develop between the robot and the participant; (d) use a rating scheme and/or emotion model that allows for simultaneous conflicting states.

5.5 Conclusions for Complexity, Voodle, and CuddleBits Behaviour Evaluation

Despite their simplicity, the CuddleBits are capable of evocative emotion display and interaction. In the above studies, we have demonstrated that participants can distinguish emotional behaviours along axes of valence and arousal[3], create complex emotional stories for the robots[3,4], and conceive of the robot as a social agent[4,5]. We have shown that it is possible to create emotional stories with the robot, and that interactors truly inter-act by becoming aligned with the robot[4]. We have further shown that the arousal and valence of a breathing-like behaviour can be determined according to some statistical measures of the behaviour, and have identified signal features that correlate with arousal and valence (i.e., com-

[3] *CuddleBits*.
[4] *Voodle*.
[5] *Complexity*.

plexity measures such as MSE for valence, and frequency/amplitude measures for arousal)[3,5]. However, we have also presented evidence that simple dimensional measures for determining an emotion state are not sufficient to capture the diversity and depth of participants' subjective experience while interacting with the robot[4]. We draw the conclusion that, although arousal is relatively consistent, valence measures are especially fragile to differences in narrative frame.

Chapter 6

Conclusions and Ongoing Work

This thesis has established the potential for simple, 1-DOF robots to engage in complex emotional interactions. If the work were to be boiled down to the a single contribution, it would be that CuddleBits can be designed to be believable social agents, despite their 'painful simplicity'. The strong caveat is that, although this sense of agency is easy to produce, it's difficult to control, due to the complexity of the human subjective experience. There are some opportunities for generality, i.e., arousal display is correlated with frequency and amplitude of the behaviour waveform; valence display with complexity measures. However, valence display needs to account for the interactor creating a narrative frame with the robot, the production of which is tied to both the interactor's experiences and the relationship they build while interacting with the robot. To reliably display valenced breathing behaviours, a designer would need to account for the impact of the narrative frame by creating some interactive context through which the robot's could be expressed and reliably interpreted.

This chapter focuses on identifying current and future work that may support valenced interactions. Work discussed here is in various stages of development in close collaboration with other SPIN researchers[1].

[1] especially Laura Cang and Mario Cimet.

6.1 Interactivity

In *Voodle*, participants continuously interacted with the robot to create emotional scenes. The robot seemed to have a much stronger ability to display emotions for those participating directly in the scene (the actor and director) than those who were emotionally removed from the scene (the observer). One interpretation of this finding could be that participants in the scene are actively suspending their disbelief, whereas the observer was not. Key to achieving this engaged emotion state is continuous interaction with the robot.

To date, the approach in the SPIN Lab has been to treat the robot's emotion display (presented here) and the robot's emotion recognition (presented in Cang 2016 [17]) as separable, even with cross-pollination between team members. A holistic approach may be necessary to properly inspect emotion recognition for both the robot and its interactor.

Such an approach would contravene current attempts at single-sample emotion measurement, i.e., when we ask participants to summarize an emotional interaction with a single rating. Emotions may only be well-defined within the context of a continuous appraisal and reappraisal of the interaction at hand with regards to a person's subjective frame (called *appraisal theory* [76]). A relational definition of emotions would require a different measurement schema than is currently used.

One approach could be to avoid any direct estimation of emotion, and instead determine the success of an emotional interaction through a goal-based approach. Imagine a study in which the goal was to *find out how the robot likes to be touched*. You might imagine the robot expressing dissatisfaction if it doesn't like to be tickled, but satisfaction if it likes to be stroked. The participant would have to reason about the robot's emotional state and determine arousal and valence continuously with direct feedback. This kind of an approach sidesteps the problems with emotion modeling by placing successful interactions as the valuation metric.

A similar idea would be to create robots with differentiable 'personalities'. Given a long enough interaction period, one could determine, e.g., whether a robot prefers to be left alone, or prefers a particular proximity. A *personality*, in this case, would be equivalent to a mapping between some set of inputs and behaviours. If an interactor were able to compare and contrast the robots by their behaviours, it would imply that some valence interpretation was consistent and successful.

6.2 Developing a Therapeutic Robot

Current work in the SPIN Lab[2] is on designing an interactive talk therapy-like session centered on the robot. The premise is to tell the robot an emotional story while touching it. The goal is to determine the emotional state of the user through their touch interactions, and, if possible, intervene through robot motion. For example, you might imagine that the robot becomes sympathetically agitated while a therapy client tells an emotionally-charged story, however, the act of calming the robot down helps them to calm themselves down.

We assume that what is necessary to this vision are (1) a system to recognize the interactor's emotion; (2) a validated set of behaviours that can be manipulated to become gradually more agitated; and (3) a mapping between the recognized emotion and the output behaviour. However, the results from *CuddleBits*, *Complexity*, and *Voodle* suggest that an emotion recognition system may not have to determine valence as much as arousal, since the interactor may interpret the robot's behaviour according to their own subjective frame. This could take the form of a simple mapping from some interactor's input to some robot motion, such as in *Voodle* where vocal input was mapped to robot position. For example, the motion of an accelerometer could be mapped to displayed arousal. This approach leaves the valence interpretation as emergent from the context of the interaction.

6.3 Can We Display Valenced Behaviours?

It would be tempting to conclude from this thesis that we cannot consistently display valenced behaviours on 1-DOF robots—that they're too dependent on the interactor's narrative frame, which is too hard to control. I would consider this viewpoint too pessimistic. We had success in displaying consistently valenced behaviours, and found correlation between signal features and valence ratings. This suggests that there are aspects of a signal that can inspire a particular interpretation of valence, but that they are easily overpowered and/or accentuated by narrative framing. Since some consistency was found (assuming no false positive), a better

[2] see Cang 2016 [17]

emotion model may also improve recognition and display relative to some specified narrative frame.

Building a believable robot must then take narrative into account. A goal of an autonomous robotic system means that the social context and the robot's situation therein must be determinable through interaction. A vision for such a system is one where the interactor can continously evaluate their actions in relation to the robot's (re)actions and build up a narrative frame over time. Results from *Voodle* and *Complexity* suggest that this is possible, since participants naturally try to create a homogenous narrative for the robot. Once a behaviour is embedded in a consistent narrative context, it is likely that it will be interpreted more consistently in terms of valence.

To motivate, consider how one learns how to interact with animals. There are some shared behaviours that map to human expressions, such as the shape and movement of a dog's eyebrows, or the height at which an animal holds its head. Many animal expressions do not have human analogues, such as a dog's wagging tail, or a cat's ears. Yet, we are able to tell what these movements mean. For some movements, we can determine valence by correlation. One can learn that a dog's wagging tail means 'happy' by observing the excitement in their eyes, noting how eagerly they approach you, or by noticing that the wagging tail accompanies a friendly pat. By contrast, we learn that a cat's wagging tail means 'angry' by watching their fur rise, their menacing growls, and by noticing that the wagging tail accompanies an unfriendly scratch. For other movements, we can determine valence by inference. If a cat leans into a stroke, one can be sure that they like it. Or, if they facilitate a stroke by making it easier to stroke them. In all cases, the context—a history with animals, a knowledge of what constitutes facilitation, repeated interactions—is what determines the valence of a particular action.

6.4 Lessons for Designing Emotional Robots

In this thesis, we have looked at the design process for simple, emotionally-expressive robots, and the generation and evaluation of their behaviours. The scope of behaviour evaluation was limited to breathing behaviours as a design challenge, and due to their emotional salience. However, we expect that many of the lessons learned here should apply to emotional machines of many configurations and capabilities.

Our findings about narrative context reinforce the understanding that humans interact emotionally with machines. Creating contexts in which machines can express emotion requires that designers facilitate the interactor seeing the machine as a social agent and creating a narrative context. Although this may not be desirable for every object—maybe we don't want a social garbage can—the interactive objects that already act in our lives socially may be enhanced by emotional programming.

For example, our robot vacuum cleaners already work within the social space of our homes, and may be much more interesting and understandable if imbued with emotionally expressive behaviours. Imagine a robot vacuum cleaner that could seem happy about its work, or frustrated at being unable to reach a spot in the room, or scared about its battery running out. Our work suggests that it may be possible to convey some of this emotional complexity with only 1-DOF motion, as long as the narrative context was accounted for in the behaviour design. Using some of our findings, we might expect a happy robot to drive smoothly, a frustrated robot to drive erratically, and a scared robot to shake erratically.

Why would we want an emotional robot vacuum cleaner? I would argue that we already do have robot vacuum cleaners that work in our human emotional space, but they're just not very good. Well-thought out emotional behaviours for a robot vacuum cleaner would be more than a cute feature, they would be critical to a good design. Humans act emotionally with their world; machines that communicate emotionally make transparent their purpose, motivations, and status. For a robot vacuum cleaner, this might equate to the robot's owner being able to understand the robot at a deeper level and make it less frustrating to be around; for a collaborative robot working in industry, such emotional insight might save someone's life. The work presented here suggests that that kind of emotional interaction is possible with just simple motion—if we can get the 'story' right.

Bibliography

[1] PhysioNet complexity and variability analysis. https://physionet.org/tutorials/cv/. Accessed: 2017-08-22. → pages 16

[2] J. Allen, L. Cang, M. Phan-Ba, A. Strang, and K. MacLean. Introducing the cuddlebot: A robot that responds to touch gestures. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*, pages 295–295. ACM, 2015. → pages 39

[3] Anki, 2017. https://anki.com/en-us/cozmo. → pages 76

[4] F. Arab, S. Paneels, M. Anastassova, S. Coeugnet, F. Le Morellec, A. Dommes, and A. Chevalier. Haptic patterns and older adults: To repeat or not to repeat? In *2015 IEEE World Haptics Conference (WHC)*, pages 248–253. IEEE, jun 2015. ISBN 978-1-4799-6624-0. doi:10.1109/WHC.2015.7177721. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=7177721. → pages 22

[5] J. N. Bailenson and N. Yee. Digital chameleons automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological science*, 16(10):814–819, 2005. → pages 21

[6] R. Banse and K. R. Scherer. Acoustic profiles in vocal emotion expression. *Journal of personality and social psychology*, 70(3):614, 1996. → pages 22

[7] J. Bates et al. The role of emotion in believable agents. *Communications of the ACM*, 37(7):122–125, 1994. → pages 24, 40

[8] A. Beck, A. Hiolle, and L. Canamero. Using perlin noise to generate emotional expressions in a robot. In *Proceedings of annual meeting of the cognitive science society (Cog Sci 2013)*, pages 1845–1850, 2013. → pages 20, 78

[9] H. R. Bernard. *Research methods in anthropology: Qualitative and quantitative approaches*. Rowman Altamira, 2011. → pages 66

[10] J. A. Bilmes, X. Li, J. Malkin, K. Kilanski, R. Wright, K. Kirchhoff, A. Subramanya, S. Harada, J. A. Landay, P. Dowden, and H. Chizeck. The Vocal Joystick: A Voice-based Human-computer Interface for Individuals with Motor Impairments. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, HLT '05, pages 995–1002, Stroudsburg, PA, USA, 2005. Association for Computational Linguistics. doi:10.3115/1220575.1220700. URL http://dx.doi.org/10.3115/1220575.1220700. → pages 22, 77

[11] S. Bloch, M. Lemeignan, and N. Aguilera-T. Specific respiratory patterns distinguish among human basic emotions. *Intl Journal of Psychophysiology*, 11(2):141–154, 1991. → pages 13, 18

[12] F. A. Boiten, N. H. Frijda, and C. J. Wientjes. Emotions and respiratory patterns: review and critical analysis. *International Journal of Psychophysiology*, 17(2):103–128, 1994. → pages 13

[13] C. Breazeal and L. Aryananda. Recognition of Affective Communicative Intent in Robot-Directed Speech. *Autonomous Robots*, 12(1):83–104, 2002. ISSN 1573-7527. doi:10.1023/A:1013215010749. URL http://dx.doi.org/10.1023/A:1013215010749. → pages 22

[14] L. Brunet, C. Megard, S. Paneels, G. Changeon, J. Lozada, M. P. Daniel, and F. Darses. "Invitation to the voyage": The design of tactile metaphors to fulfill occasional travelers' needs in transportation networks. In *2013 World Haptics Conference (WHC)*, pages 259–264. IEEE, apr 2013. ISBN 978-1-4799-0088-6. doi:10.1109/WHC.2013.6548418. URL http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=6548418. → pages 22

[15] P. Bucci, X. L. Cang, M. Chun, D. Marino, O. Schneider, H. Seifi, and K. MacLean. CuddleBits: an iterative prototyping platform for complex haptic display. In *EuroHaptics '16 Demos*, 2016. → pages 13

[16] P. Bucci, L. Cang, S. Valair, D. Marino, L. Tseng, M. Jung, J. Rantala, O. Schneider, and K. MacLean. Sketching cuddlebits: Coupled prototyping of body and behaviour for an affective robot pet. *To appear in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI) 2017*, 2017. → pages iv, 49, 77

[17] L. Cang. Towards an emotionally communicative robot: feature analysis for multimodal support of affective touch recognition. Master's thesis, University of British Columbia, 2016. → pages 26, 91, 92

[18] L. Cang, P. Bucci, and K. E. MacLean. Cuddlebits: Friendly, low-cost furballs that respond to touch. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pages 365–366. ACM, 2015. doi:10.1145/2818346.2823293. → pages 13

[19] X. L. Cang, P. Bucci, A. Strang, J. Allen, K. MacLean, and H. Liu. Different strokes and different folks: Economical dynamic surface sensing and affect-related touch recognition. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, pages 147–154. ACM, 2015. → pages 25

[20] N. Chomsky and M. Halle. *The sound pattern of English.* New York: Harper and Row, 1968. → pages 46

[21] H. Choset, K. M. Lynch, S. Hutchinson, G. Kantor, W. Burgard, L. E. Kavraki, and S. Thrun. *Principles of Robot Motion: Theory, Algorithms, and Implementations*. The MIT Press, 2005. → pages 20

[22] M. Chung, E. Rombokas, Q. An, Y. Matsuoka, and J. Bilmes. Continuous vocalization control of a full-scale assistive robot. In *2012 4th IEEE RAS EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*, pages 1464–1469, jun 2012. doi:10.1109/BioRob.2012.6290664. → pages 22

[23] B. Clark, O. S. Schneider, K. E. MacLean, and H. Z. Tan. Predictable and distinguishable morphing of vibrotactile rhythm. → pages 87

[24] J. Corbin and A. Strauss. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*. Sage Publications, Inc., 3 edition, 2008. → pages 67

[25] M. Costa, A. L. Goldberger, and C.-K. Peng. Multiscale entropy analysis of complex physiologic time series. *Physical review letters*, 89(6):068102, 2002. → pages 17

[26] M. D. Costa, C.-K. Peng, and A. L. Goldberger. Multiscale analysis of heart rate dynamics: entropy and time irreversibility measures. *Cardiovascular Engineering*, 8(2):88–93, 2008. → pages 17, 80

[27] J. Q. Dawson, O. S. Schneider, J. Ferstay, D. Toker, J. Link, S. Haddad, and K. MacLean. It's alive!: exploring the design space of a gesturing phone. In *Proc of Graphics Interface 2013*, pages 205–212. Canadian Information Processing Society, 2013. → pages 54

[28] C. M. de Melo, P. Kenny, and J. Gratch. Real-time expression of affect through respiration. *Computer Animation and Virtual Worlds*, 21(3-4): 225–234, 2010. → pages 13, 18

[29] F. De Saussure, W. Baskin, and P. Meisel. *Course in general linguistics*. Columbia University Press, 2011. → pages 21

[30] L. Feldman Barrett and J. A. Russell. Independence and bipolarity in the structure of current affect. *Journal of personality and social psychology*, 74 (4):967, 1998. → pages 15

[31] T. Fong, I. Nourbakhsh, and K. Dautenhahn. A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42(3):143–166, 2003. → pages 12

[32] J. Forsslund, M. Yip, and E.-L. Sallnäs. Woodenhaptics: A starting kit for crafting force-reflecting spatial haptic devices. In *Proceedings of the Ninth International Conference on Tangible, Embedded, and Embodied Interaction*, pages 133–140. ACM, 2015. → pages 19

[33] S. Garrod and M. J. Pickering. Why is conversation so easy? *Trends in cognitive sciences*, 8(1):8–11, 2004. → pages 21, 76

[34] M. Goto, K. Kitayama, K. Itou, and T. Kobayashi. Speech Spotter: On-demand Speech Recognition. In *in Human-Human Conversation on the Telephone or in Face-to-Face Situations. Proc. ICSLP'04*, pages 1533–1536, 2004. → pages 22

[35] J. Gratch, A. Okhmatovskaia, F. Lamothe, S. Marsella, M. Morales, R. J. van der Werf, and L.-P. Morency. Virtual rapport. In *International Workshop on Intelligent Virtual Agents*, pages 14–27. Springer, 2006. → pages 21

[36] J. Gray, G. Hoffman, S. O. Adalgeirsson, M. Berlin, and C. Breazeal. Expressive, interactive robots: Tools, techniques, and insights based on collaborations. In *Human Robot Interaction (HRI) 2010 Workshop: What do collaborations with the arts have to say about HRI*, 2010. → pages 20, 76

[37] S. Harada, J. A. Landay, J. Malkin, X. Li, and J. A. Bilmes. The Vocal Joystick:: Evaluation of Voice-based Cursor Control Techniques. In *Proceedings of the 8th International ACM SIGACCESS Conference on Computers and Accessibility*, Assets '06, pages 197–204, New York, NY, USA, 2006. ACM. ISBN 1-59593-290-9. doi:10.1145/1168987.1169021. URL http://doi.acm.org/10.1145/1168987.1169021. → pages 77

[38] G. Hoffman. On stage: robots as performers. In *RSS 2011 Workshop on Human-Robot Interaction: Perspectives and Contributions to Robotics from the Human Sciences. Los Angeles, CA*, volume 1, 2011. → pages 21

[39] G. Hoffman and W. Ju. Designing robots with movement in mind. *Journal of Human-Robot Interaction*, 3(1):89–122, 2014. → pages 20

[40] G. Hoffman and W. Ju. Designing robots with movement in mind. *Journal of Human-Robot Interaction*, 3(1):89–122, 2014. → pages 19, 77

[41] B. House, J. Malkin, and J. Bilmes. The VoiceBot: A Voice Controlled Robot Arm. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 183–192, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-246-7. doi:10.1145/1518701.1518731. URL http://doi.acm.org/10.1145/1518701.1518731. → pages 22

[42] T. Igarashi and J. F. Hughes. Voice As Sound: Using Non-verbal Voice Input for Interactive Control. In *Proceedings of the 14th Annual ACM Symposium on User Interface Software and Technology*, UIST '01, pages 155–156, New York, NY, USA, 2001. ACM. ISBN 1-58113-438-X. doi:10.1145/502348.502372. URL http://doi.acm.org/10.1145/502348.502372. → pages 22

[43] Johnny-Five, 2017. http://johnny-five.io. → pages 47

[44] K. J. Kitto. *Modelling and generating complex emergent behaviour*. Flinders University, School of Chemistry, Physics and Earth Sciences., 2006. → pages 16

[45] K. Krippendorff. Computing krippendorff's alpha reliability. *Departmental papers (ASC)*, page 43, 2007. → pages 81

[46] J. L. Lakin, V. E. Jefferis, C. M. Cheng, and T. L. Chartrand. The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. *Journal of nonverbal behavior*, 27(3):145–162, 2003. → pages 21

[47] G. Lakoff and M. Johnson. *Metaphors we live by*. 2003. → pages 70

[48] Y.-K. Lim, E. Stolterman, and J. Tenenberg. The anatomy of prototypes: Prototypes as filters, prototypes as manifestations of design ideas. *ACM TOCHI*, 15(2):7, 2008. → pages 19

[49] R. Maatman, J. Gratch, and S. Marsella. Natural behavior of a listening agent. In *International Workshop on Intelligent Virtual Agents*, pages 25–36. Springer, 2005. → pages 21

[50] M. M. Marin and H. Leder. Examining complexity across domains: relating subjective and objective measures of affective environmental scenes, paintings and music. *PloS One*, 8(8):e72412, 2013. → pages 16

[51] D. Marino, P. Bucci, O. S. Schneider, and K. E. MacLean. Voodle: Vocal doodling to sketch affective robot motion. In *Proceedings of the 2017 Conference on Designing Interactive Systems*, pages 753–765. ACM, 2017. → pages iv

[52] L. Mathiassen, T. Seewaldt, and J. Stage. Prototyping and specifying: principles and practices of a mixed approach. *Scandinavian Journal of Information Systems*, 7(1):4, 1995. → pages 19

[53] D. H. McFarland. Respiratory markers of conversational interaction. *Journal of Speech, Language, and Hearing Research*, 44(1):128–143, 2001. → pages 21

[54] A. Moon, C. A. Parker, E. A. Croft, and H. M. Van der Loos. Did you see it hesitate?–empirically grounded design of hesitation trajectories for collaborative robots. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1994–1999. IEEE, 2011. → pages 20

[55] C. Moussette. Sketching in Hardware and Building Interaction Design : tools , toolkits and an attitude for Interaction Designers. In *Processing*, 2010. URL https://public.me.com/intuitive. → pages 19

[56] C. Moussette. Simple haptics: Sketching perspectives for the design of haptic interactions. *Umeå Universitet*, 2012. → pages 19

[57] C. Moussette, S. Kuenen, and A. Israr. Designing haptics. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction - TEI '12*, page 351, New York, New York, USA, feb 2012. ACM Press. ISBN 9781450311748. doi:10.1145/2148131.2148215. URL http://dl.acm.org/citation.cfm?id=2148131.2148215. → pages 7

[58] S. Oliver and M. Karon. Haptic jazz: Collaborative touch with the haptic instrument. In *IEEE Haptics Symposium*, 2014. → pages 20

[59] J. S. Pardo. On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America*, 119(4):2382–2393, 2006. → pages 21

[60] J. Pearson, J. Hu, H. P. Branigan, M. J. Pickering, and C. I. Nass. Adaptive language behavior in hci: how expectations and beliefs about a system affect users' word choice. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 1177–1180. ACM, 2006. → pages 21

[61] C. Peirce. *The philosophical writings of Peirce*, pages 98–119. New York: Dover, 1955. → pages 22

[62] M. Perlman and A. A. Cain. Iconicity in vocalization, comparisons with gesture, and implications for theories on the evolution of language. *Gesture*, 14(3):320–350, 2014. → pages 22

[63] P. Perniss and G. Vigliocco. The bridge of iconicity: from a world of experience to the experience of language. *Phil. Trans. R. Soc. B*, 369(1651): 20130300, 2014. → pages 22

[64] P. Rainville, A. Bechara, N. Naqvi, and A. R. Damasio. Basic emotions are associated with distinct patterns of cardiorespiratory activity. *International journal of psychophysiology*, 61(1):5–18, 2006. → pages 13

[65] J. Rantala, K. Salminen, R. Raisamo, and V. Surakka. Touch gestures in communicating emotional intention via vibrotactile stimulation. *Intl J Human-Computer Studies*, 71(6):679–690, 2013. → pages 54

[66] React, 2017. https://facebook.github.io/react. → pages 47

[67] T. Ribeiro and A. Paiva. The illusion of robotic life: principles and practices of animation for robots. In *Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction*, pages 383–390. ACM, 2012. → pages 20

[68] J. S. Richman and J. R. Moorman. Physiological time-series analysis using approximate entropy and sample entropy. *American Journal of Physiology-Heart and Circulatory Physiology*, 278(6):H2039–H2049, 2000. → pages 17

[69] R. Rose, M. Scheutz, and P. Schermerhorn. Towards a conceptual and methodological framework for determining robot believability. *Interaction Studies*, 11(2):314–335, 2010. → pages 23

[70] R. Rummer, J. Schweppe, R. Schlegelmilch, and M. Grice. Mood is linked to vowel type: The role of articulatory movements. *Emotion*, 14(2):246, 2014. → pages 22

[71] J. A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161, 1980. → pages 15

[72] J. A. Russell, A. Weiss, and G. A. Mendelsohn. Affect grid: a single-item scale of pleasure and arousal. *Journal of personality and social psychology*, 57(3):493, 1989. → pages 15, 52

[73] G. W. Ryan and H. R. Bernard. Techniques to Identify Themes. *Field Methods*, 15(1):85–109, feb 2003. ISSN 00000000. doi:10.1177/1525822X02239569. URL http://fmx.sagepub.com/cgi/doi/10.1177/1525822X02239569. → pages 67

[74] D. Sakamoto, T. Komatsu, and T. Igarashi. Voice Augmented Manipulation: Using Paralinguistic Information to Manipulate Mobile Devices. In *Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services*, MobileHCI '13, pages 69–78, New York, NY, USA, 2013. ACM. ISBN 978-1-4503-2273-7. doi:10.1145/2493190.2493244. URL http://doi.acm.org/10.1145/2493190.2493244. → pages 22

[75] J. Saldien, K. Goris, S. Yilmazyildiz, W. Verhelst, and D. Lefeber. On the design of the huggable robot probo. *Journal of Physical Agents*, 2(2):3–12, 2008. → pages 20

[76] K. R. Scherer. Appraisal theory. 1999. → pages 91

[77] O. S. Schneider and K. E. MacLean. Improvising design with a haptic instrument. In *2014 IEEE Haptics Symposium (HAPTICS)*, pages 327–332. IEEE, 2014. → pages 22, 78

[78] O. S. Schneider and K. E. MacLean. Studying Design Process and Example Use with Macaron, a Web-based Vibrotactile Effect Editor. In *HAPTICS '16: Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems*, 2016. → pages 48

[79] O. S. Schneider and K. E. MacLean. Studying design process and example use with macaron, a web-based vibrotactile effect editor. In *2016 IEEE Haptics Symposium (HAPTICS)*, pages 52–58. IEEE, 2016. → pages 8, 20, 39

[80] O. S. Schneider, H. Seifi, S. Kashani, M. Chun, and K. E. MacLean. HapTurk: Crowdsourcing Affective Ratings for Vibrotactile Icons. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI) '16*, pages 3248–3260, New York, New York, USA, may 2016. ACM Press. ISBN 9781450333627. doi:10.1145/2858036.2858279. URL http://dl.acm.org/citation.cfm?id=2858036.2858279. → pages 43

[81] Y. S. Sefidgar, K. E. MacLean, S. Yohanan, H. M. Van der Loos, E. A. Croft, and E. J. Garland. Design and evaluation of a touch-centered calming interaction with a social robot. *IEEE Transactions on Affective Computing*, 7 (2):108–121, 2016. → pages 13

[82] H. Seifi, K. Zhang, and K. E. MacLean. Vibviz: Organizing, visualizing and navigating vibration libraries. In *World Haptics Conference (WHC), 2015 IEEE*, pages 254–259. IEEE, 2015. → pages 43

[83] M. Shaver and K. Maclean. The twiddler: A haptic teaching tool for low-cost communication and mechanical design. Master's thesis, 2003. → pages 39

[84] H. Shintel, H. C. Nusbaum, and A. Okrent. Analog acoustic expression in speech communication. *Journal of Memory and Language*, 55(2):167–177, 2006. → pages 22

[85] J. Six, O. Cornelis, and M. Leman. TarsosDSP, a Real-Time Audio Processing Framework in Java. In *Proceedings of the 53rd AES Conference (AES 53rd)*, 2014. → pages 47

[86] A. Strauss and J. Corbin. *Basics of qualitative research: Techniques and procedures for developing grounded theory* . Sage Publications, Inc, 1998. → pages 65

[87] L. Takayama, D. Dooley, and W. Ju. Expressing thought: improving robot readability with animation principles. In *Proceedings of the 6th international conference on Human-robot interaction*, pages 69–76. ACM, 2011. → pages 20

[88] K. Wada and T. Shibata. Living with seal robotsits sociopsychological and physiological influences on the elderly at a care house. *IEEE Transactions on Robotics*, 23(5):972–980, 2007. → pages 12

[89] K. Wada, T. Shibata, T. Asada, and T. Musha. Robot therapy for prevention of dementia at home. *Journal of Robotics and Mechatronics*, 19(6):691, 2007. → pages 12

[90] R. M. Warner. Coordinated cycles in behavior and physiology during face-to-face social interactions. *Sage Publications, Inc*, 1996. → pages 21

[91] J. Watanabe and M. Sakamoto. Comparison between onomatopoeias and adjectives for evaluating tactile sensations. In *Proceedings of the 6th International Conference of Soft Computing and Intelligent Systems and the 13th International Symposium on Advanced Intelligent Systems (SCIS-ISIS 2012)*, pages 2346–2348, 2012. → pages 22, 78

[92] D. Watson, L. A. Clark, and A. Tellegen. Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6):1063–1070, 1988. doi:10.1037/0022-3514.54.6.1063. URL http://doi.apa.org/getdoi.cfm?doi=10.1037/0022-3514.54.6.1063. → pages 15

[93] D. Watson, L. A. Clark, and A. Tellegen. Development and validation of brief measures of positive and negative affect: the panas scales. *Journal of personality and social psychology*, 54(6):1063, 1988. → pages 47

[94] A. C. Weidman, C. M. Steckler, and J. L. Tracy. The jingle and jangle of emotion assessment: Imprecise measurement, casual scale usage, and conceptual fuzziness in emotion research. 2016. → pages 14

[95] S. Yohanan and K. MacLean. A tool to study affective touch. In *CHI'09 Extended Abstracts*, pages 4153–4158. ACM, 2009. → pages 12

[96] S. Yohanan and K. E. MacLean. Design and assessment of the haptic creature's affect display. In *Proceedings of the 6th international conference on Human-robot interaction*, pages 473–480. ACM, 2011. → pages 12

[97] S. Yohanan and K. E. MacLean. The role of affective touch in human-robot interaction: Human intent and expectations in touching the haptic creature. *Intl J Social Robotics*, 4(2):163–180, 2012. → pages 54

Appendix A

Supporting Materials

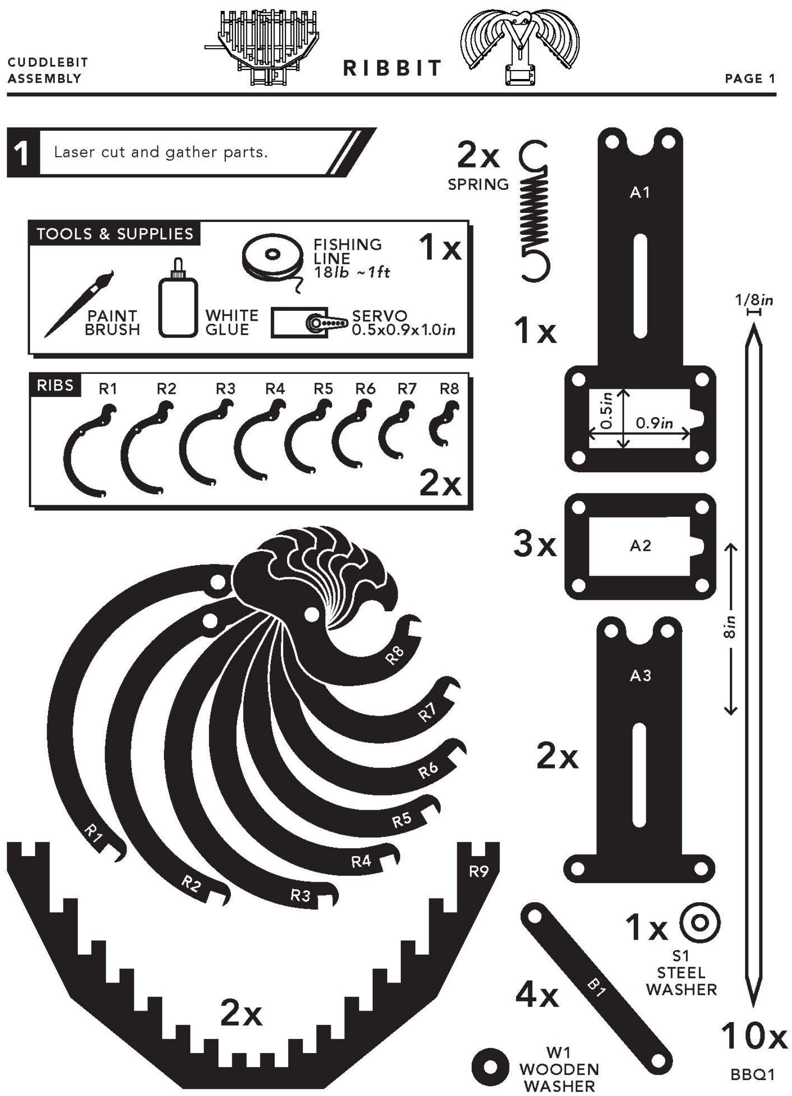


Figure A.1: RibBit assembly instructions, page 1.

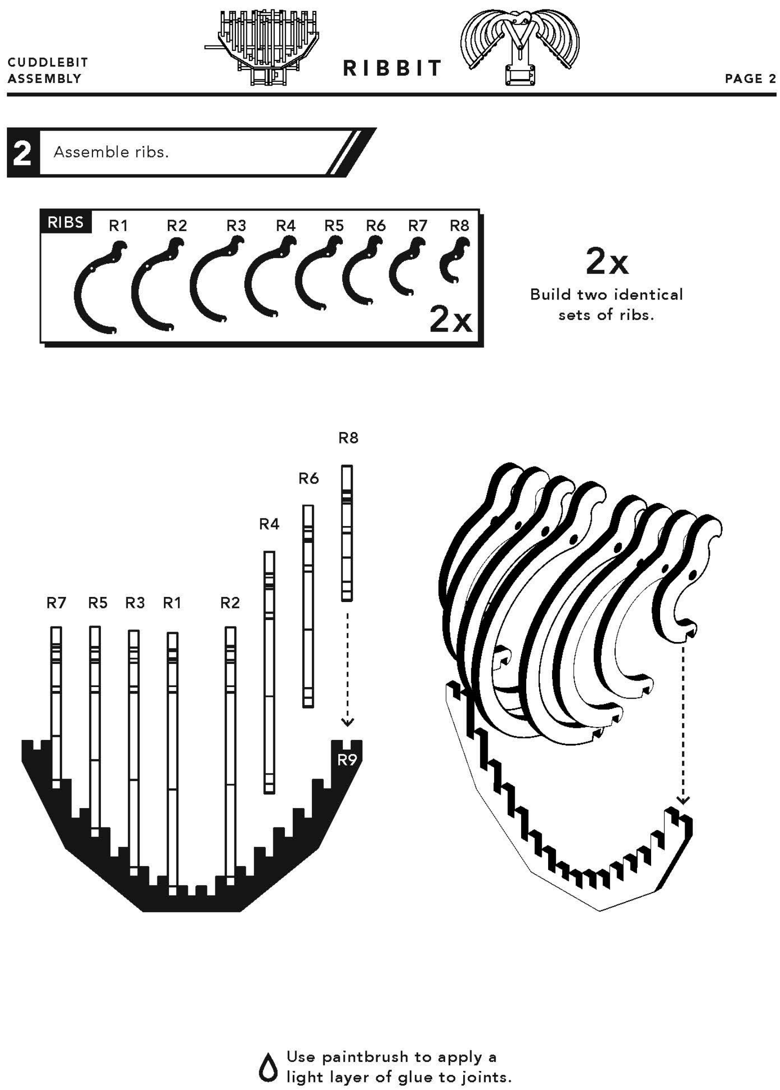


Figure A.2: RibBit assembly instructions, page 2.

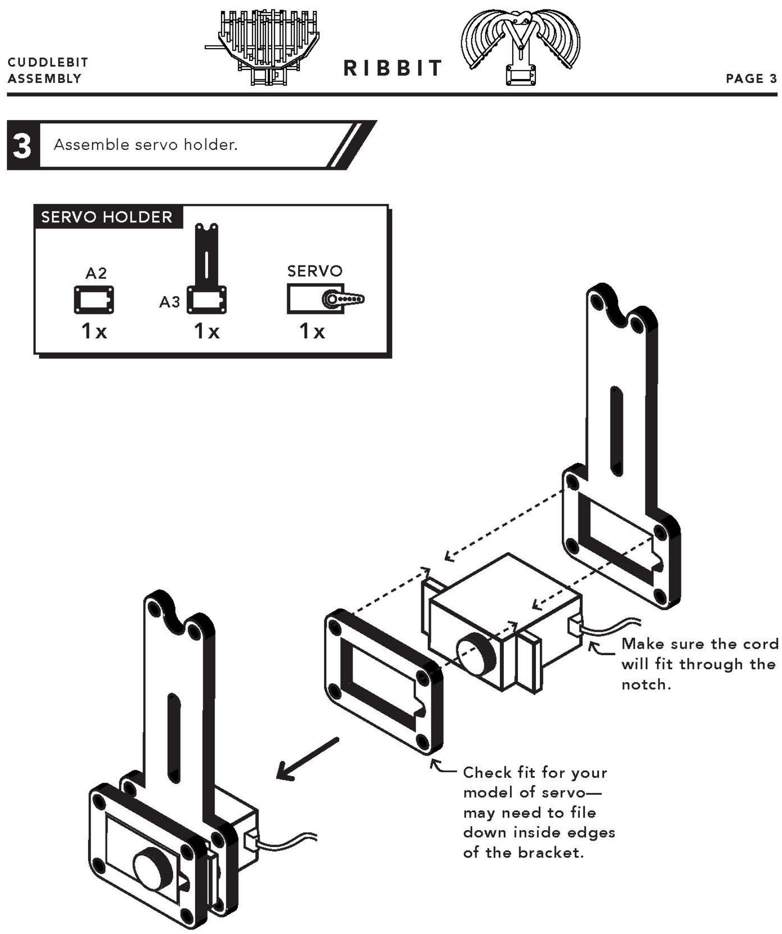


Figure A.3: RibBit assembly instructions, page 3.

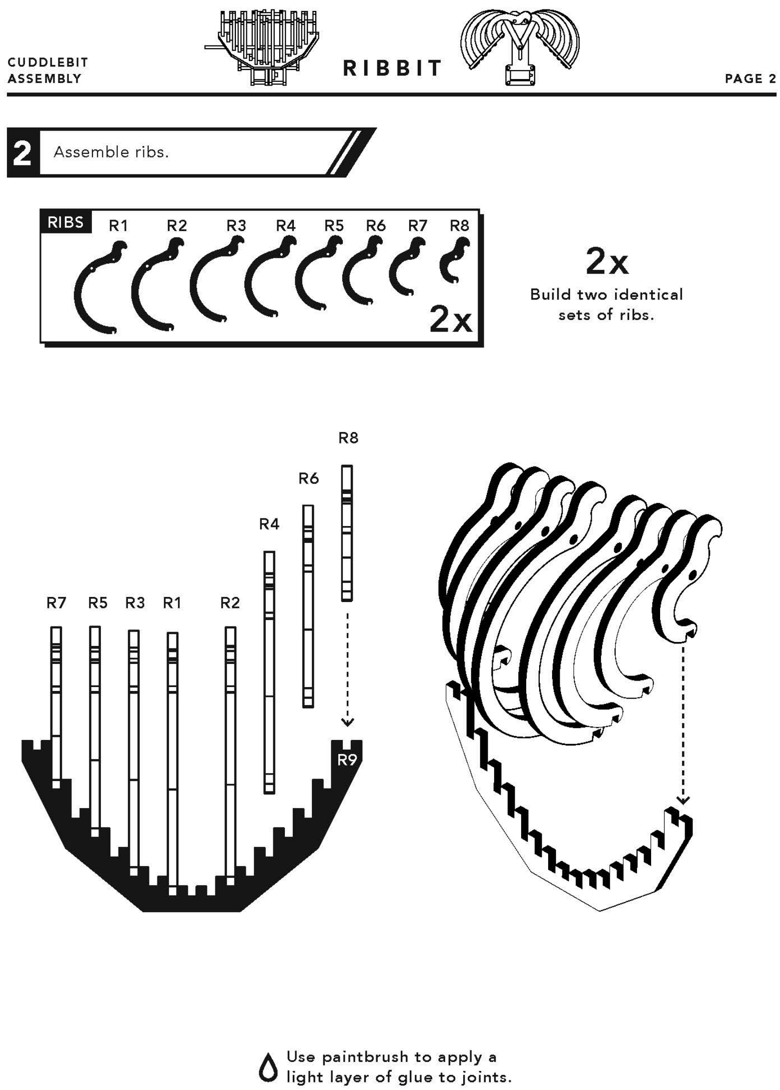


Figure A.4: RibBit assembly instructions, page 2.

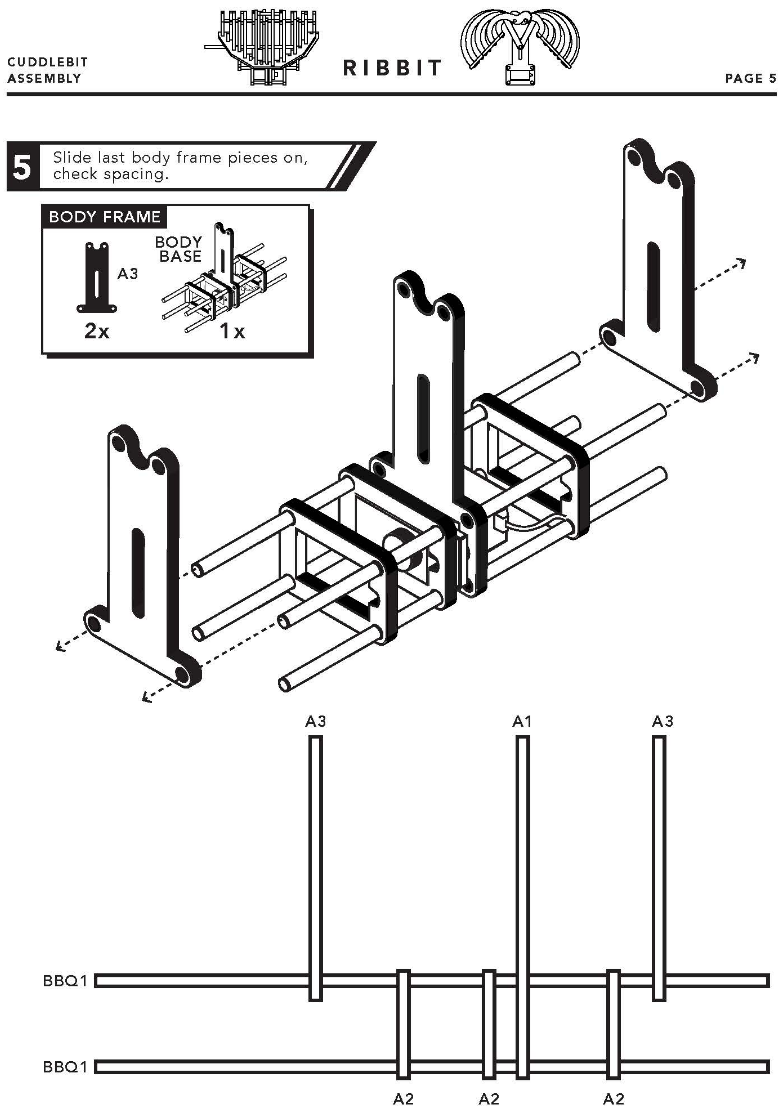


Figure A.5: RibBit assembly instructions, page 5.

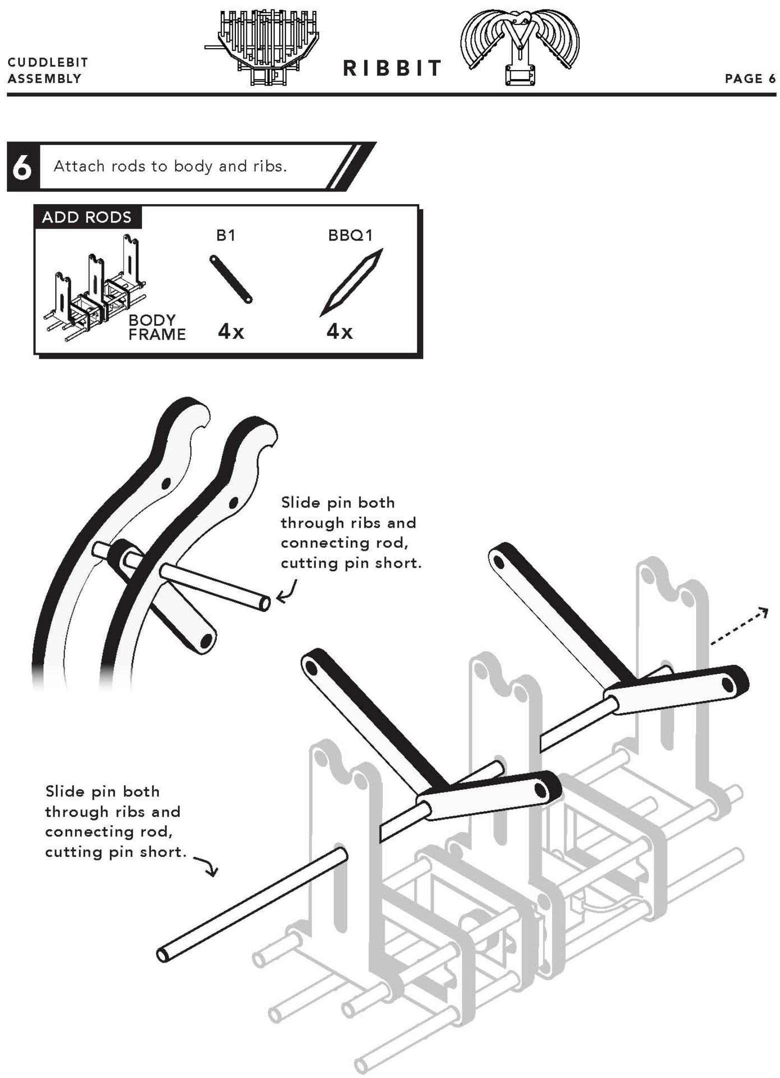


Figure A.6: RibBit assembly instructions, page 6.

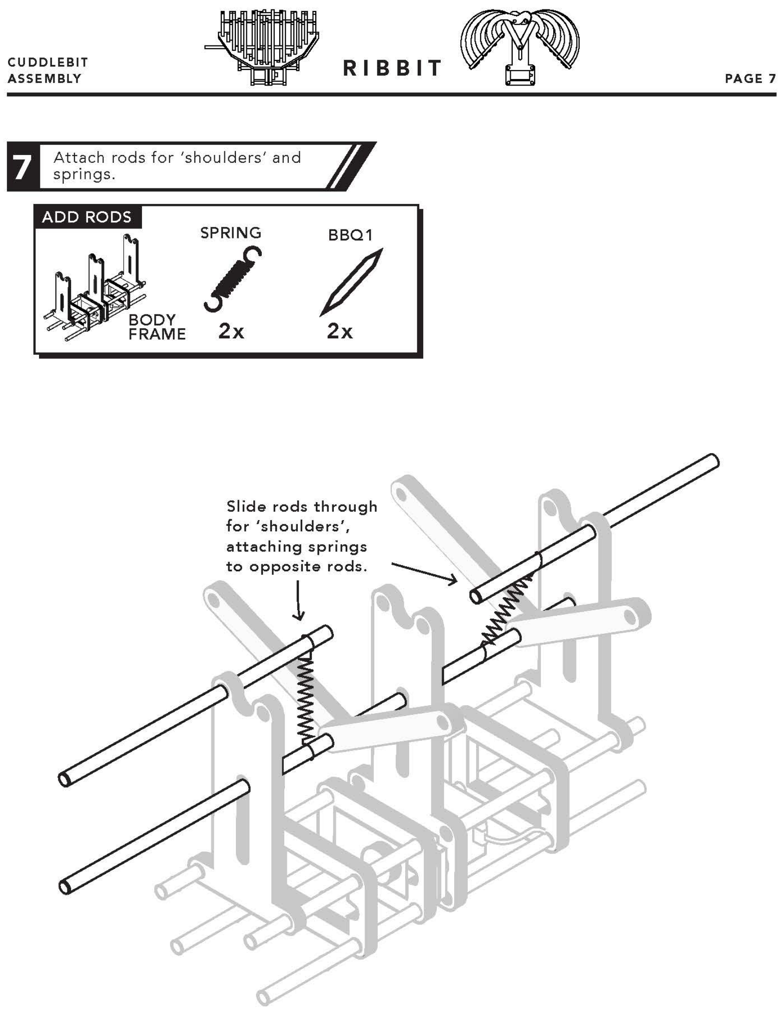


Figure A.7: RibBit assembly instructions, page 7.

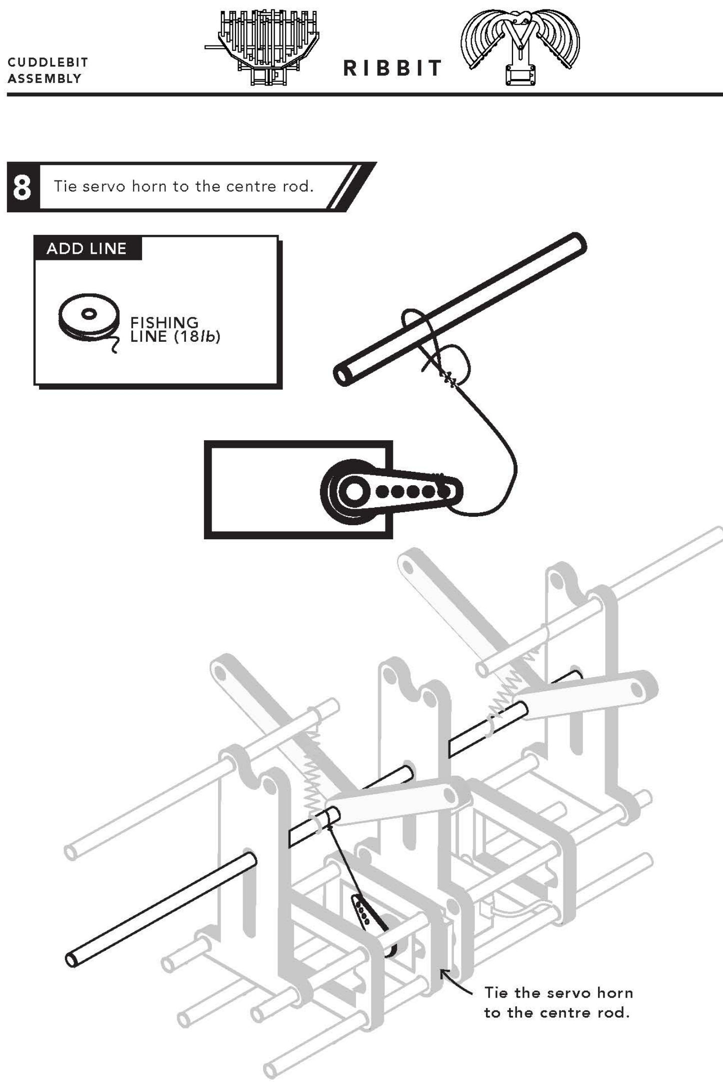


Figure A.8: RibBit assembly instructions, page 8.

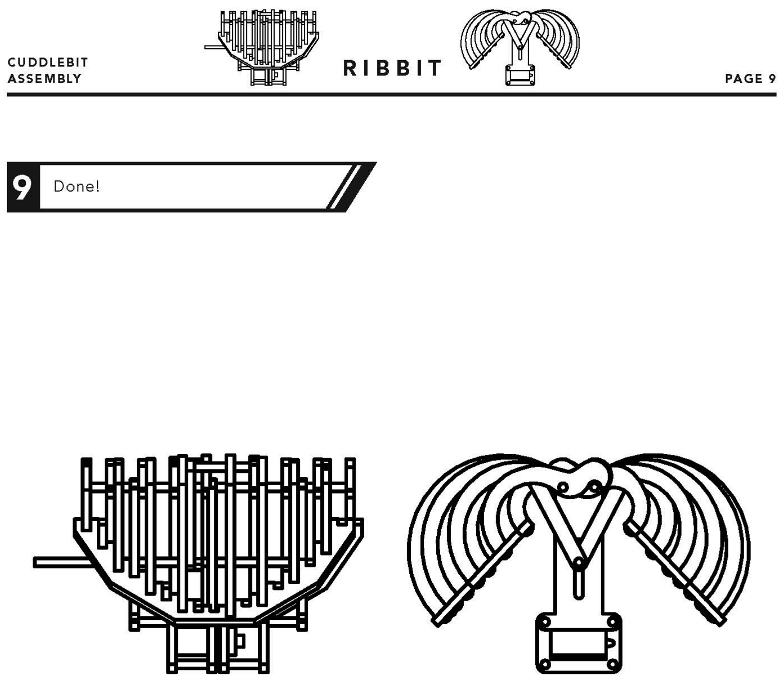


Figure A.9: RibBit assembly instructions, page 9.

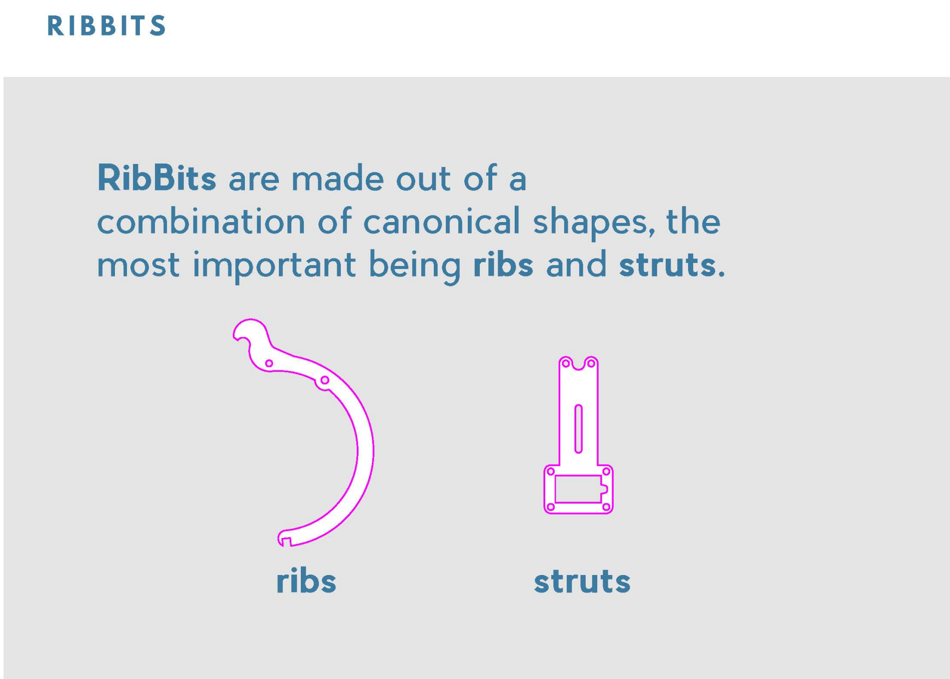


Figure A.10: RibBit design system explainer, page 1.

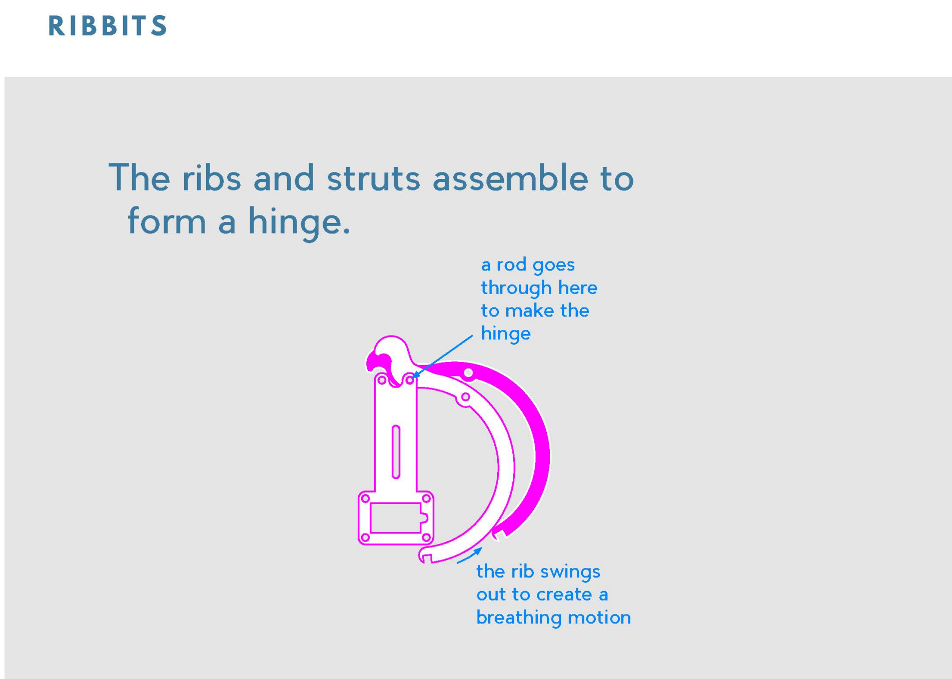


Figure A.11: RibBit design system explainer, page 2.

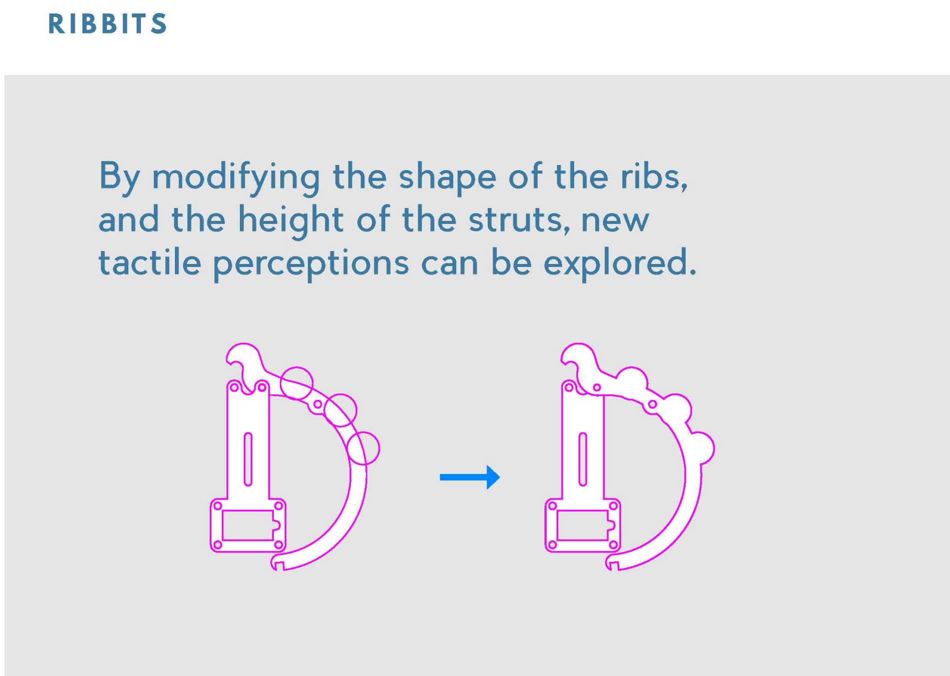


Figure A.12: RibBit design system explainer, page 3.

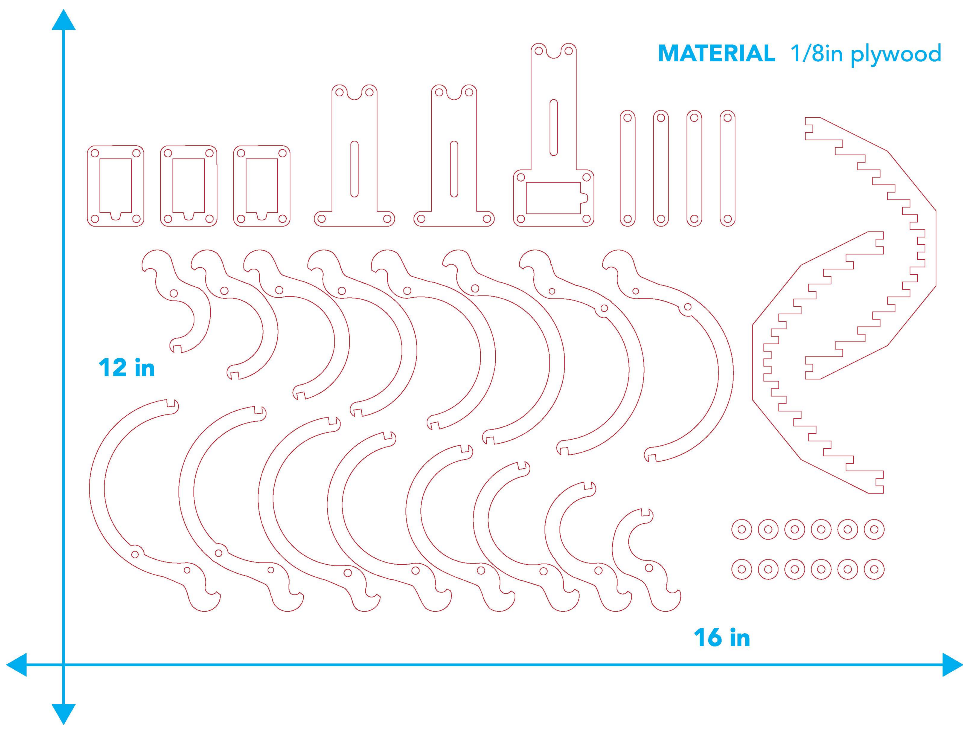


Figure A.13: Lasercutting files for the RibBit.

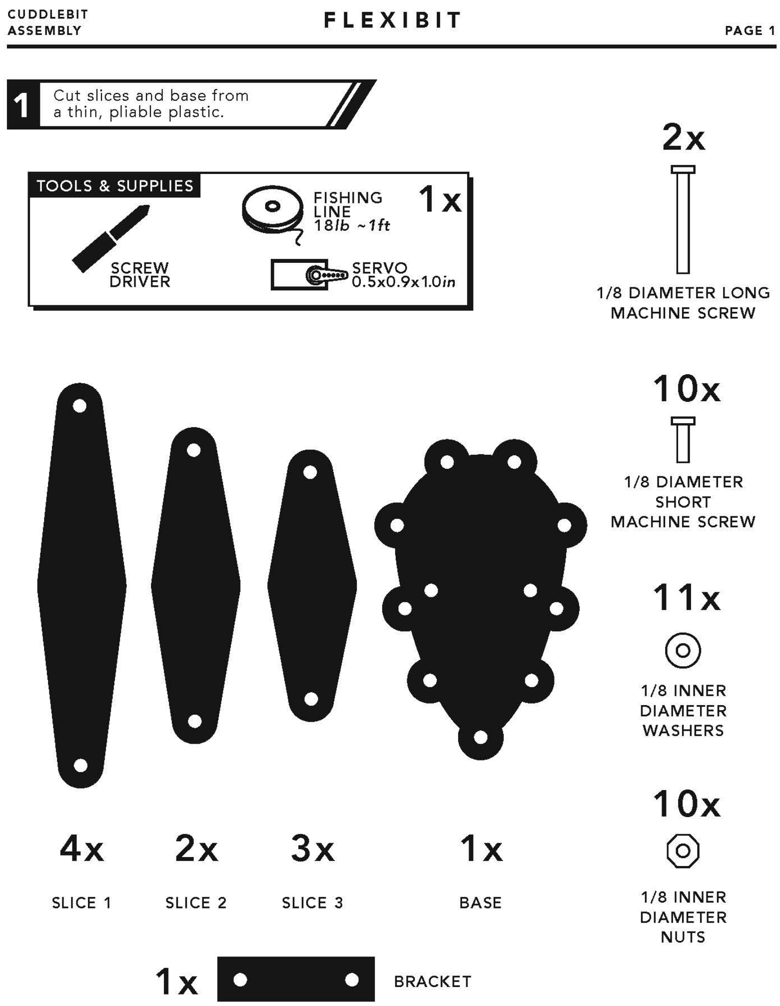


Figure A.14: FlexiBit assembly instructions, page 1.

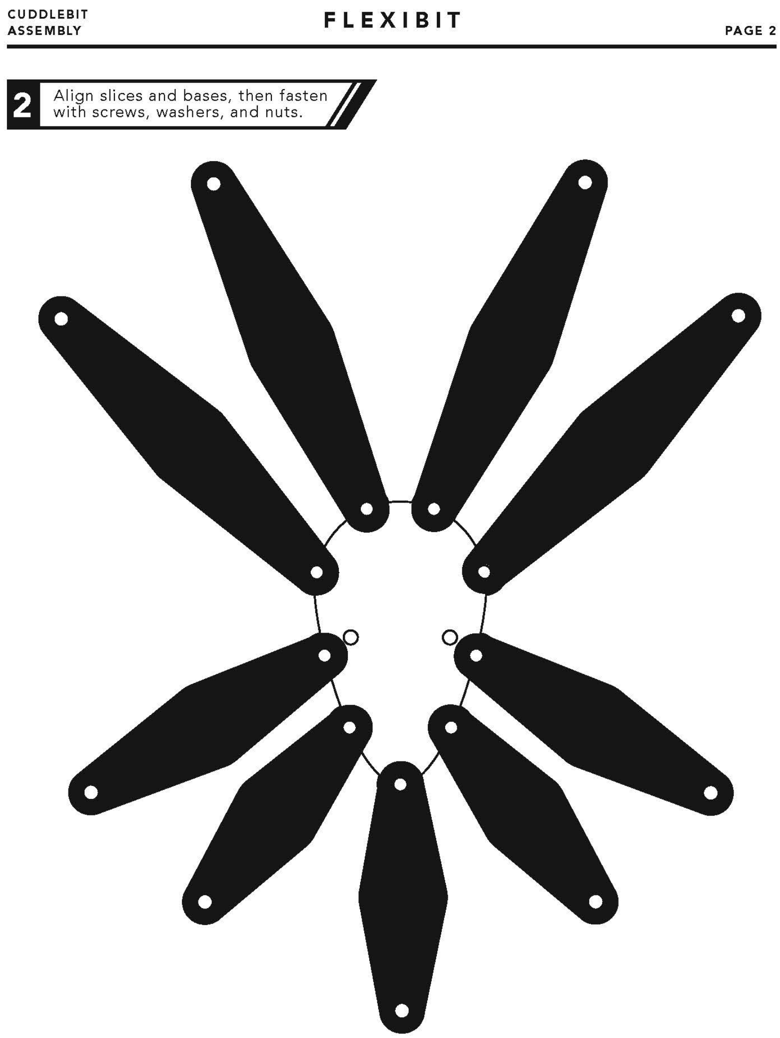


Figure A.15: FlexiBit assembly instructions, page 2.

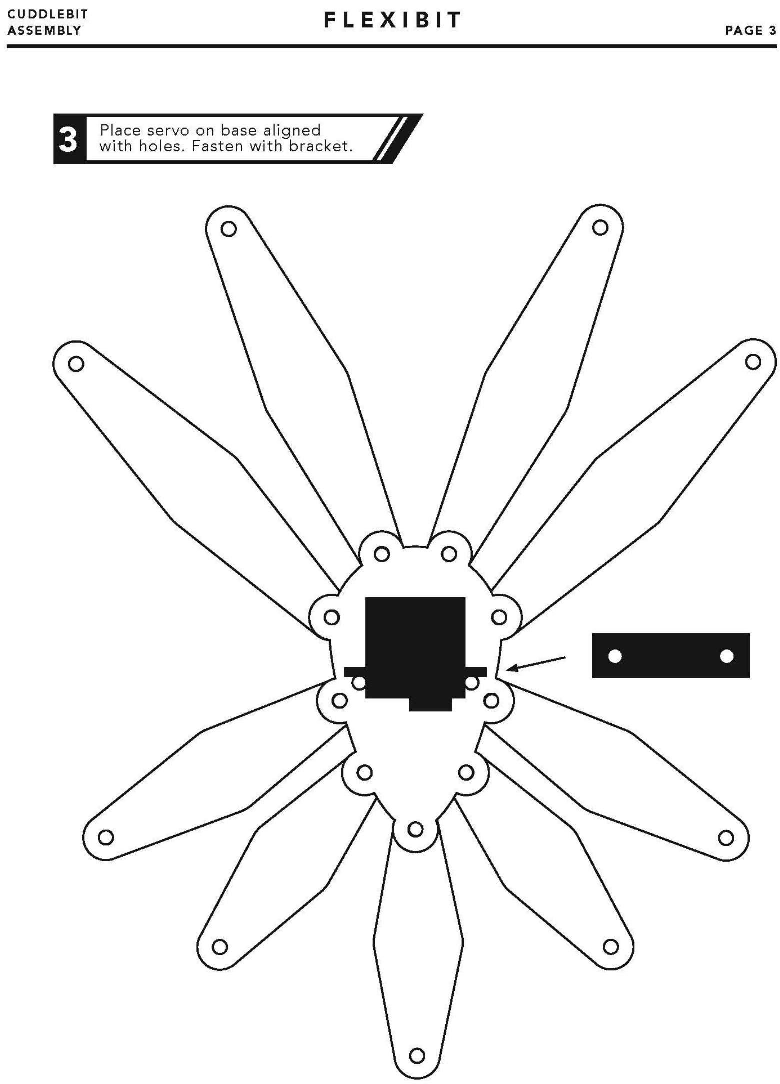


Figure A.16: FlexiBit assembly instructions, page 3.

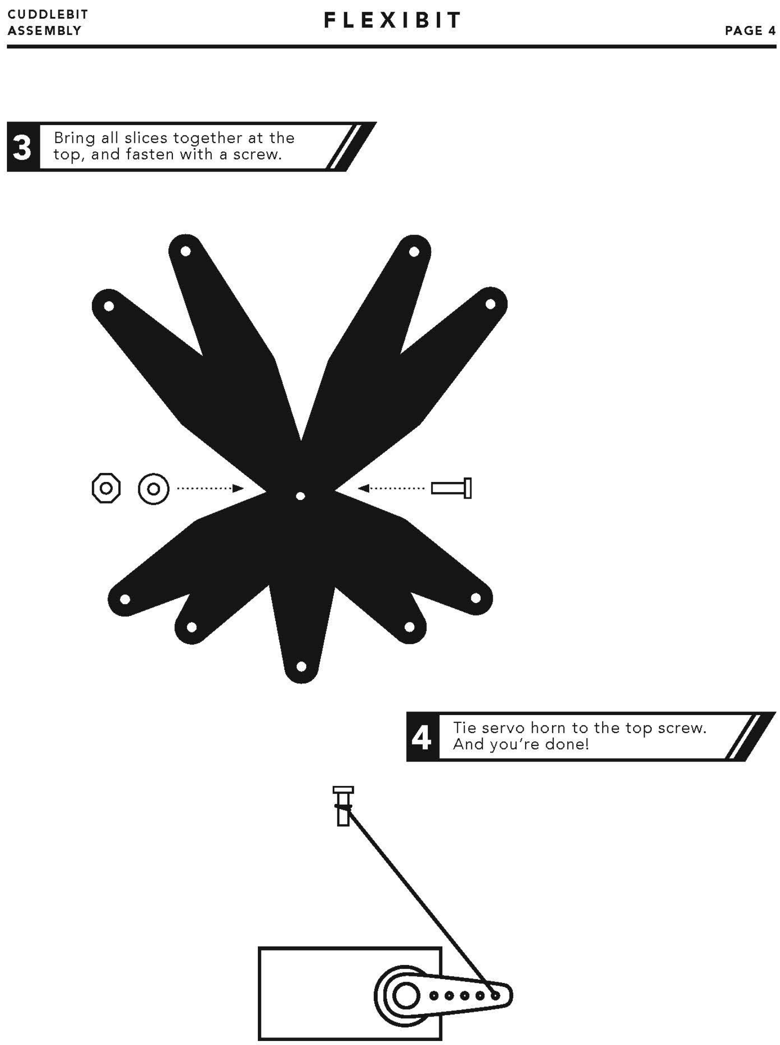


Figure A.17: FlexiBit assembly instructions, page 4.

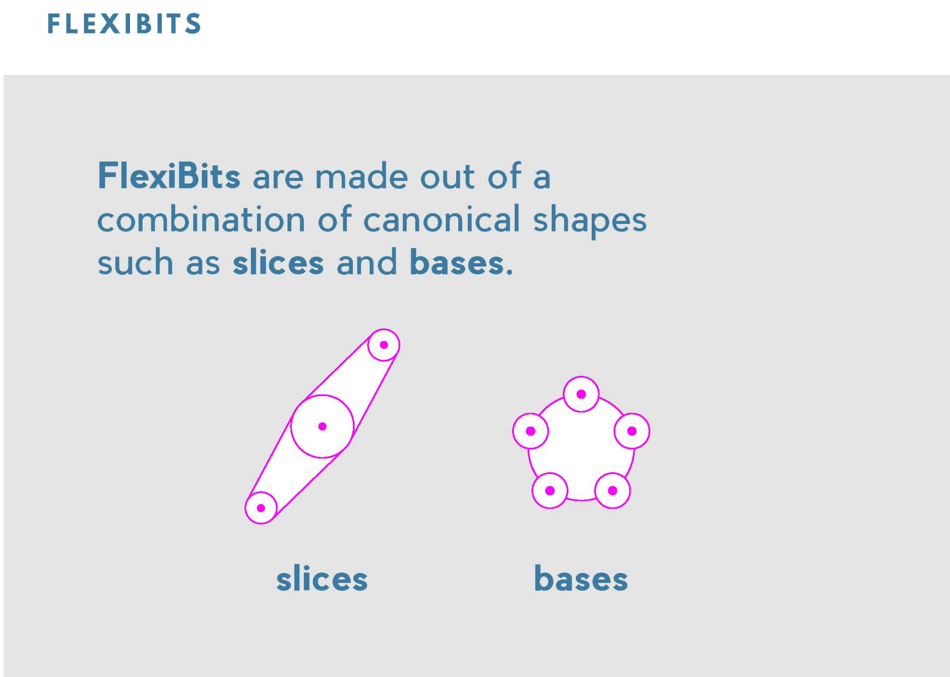


Figure A.18: FlexiBit design system explainer, page 1.

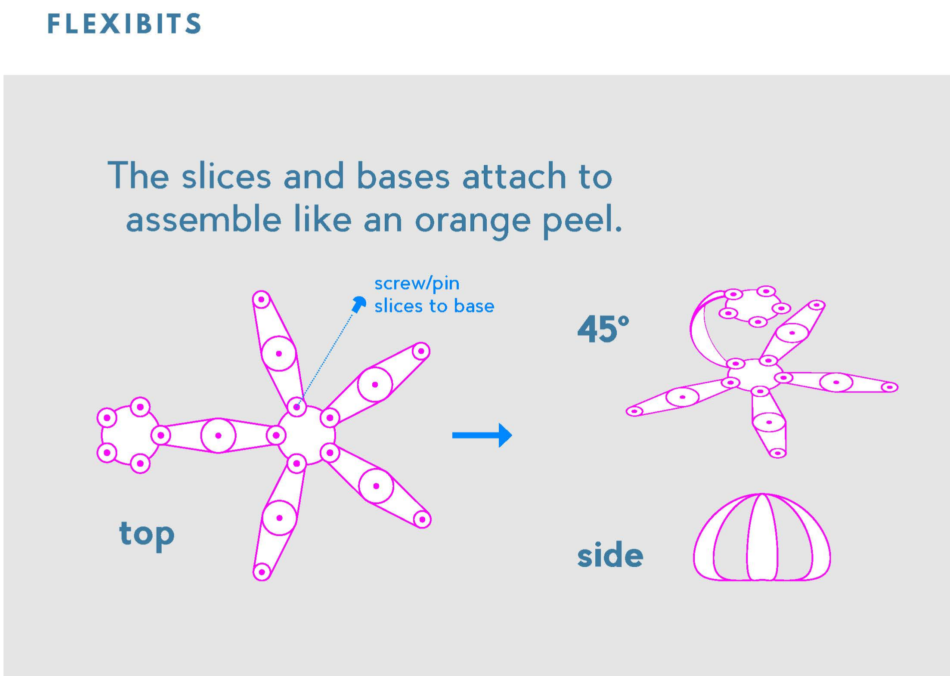


Figure A.19: FlexiBit design system explainer, page 2.

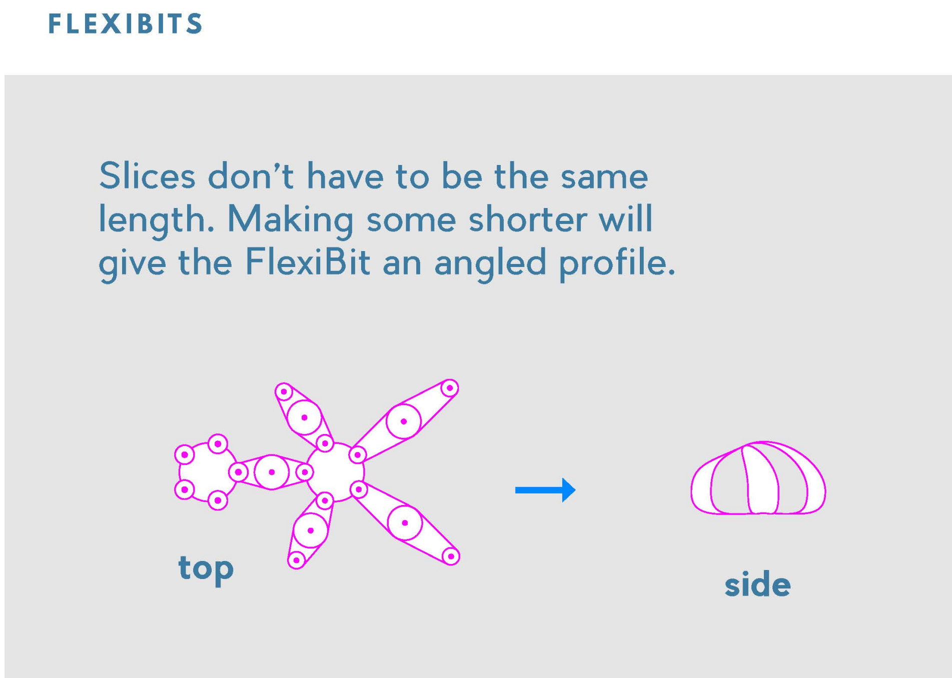


Figure A.20: FlexiBit design system explainer, page 3.

FLEXIBITS

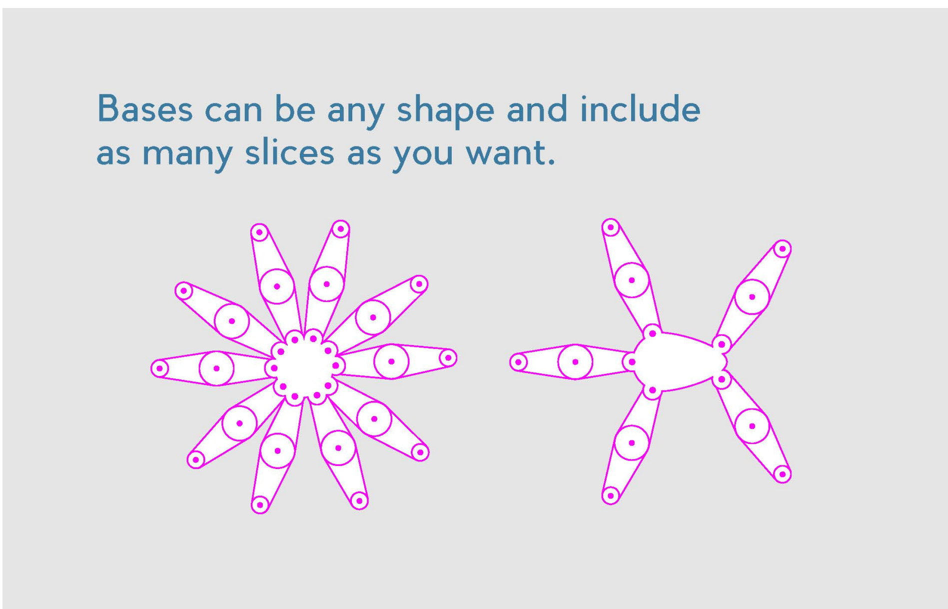


Figure A.21: FlexiBit design system explainer, page 4.

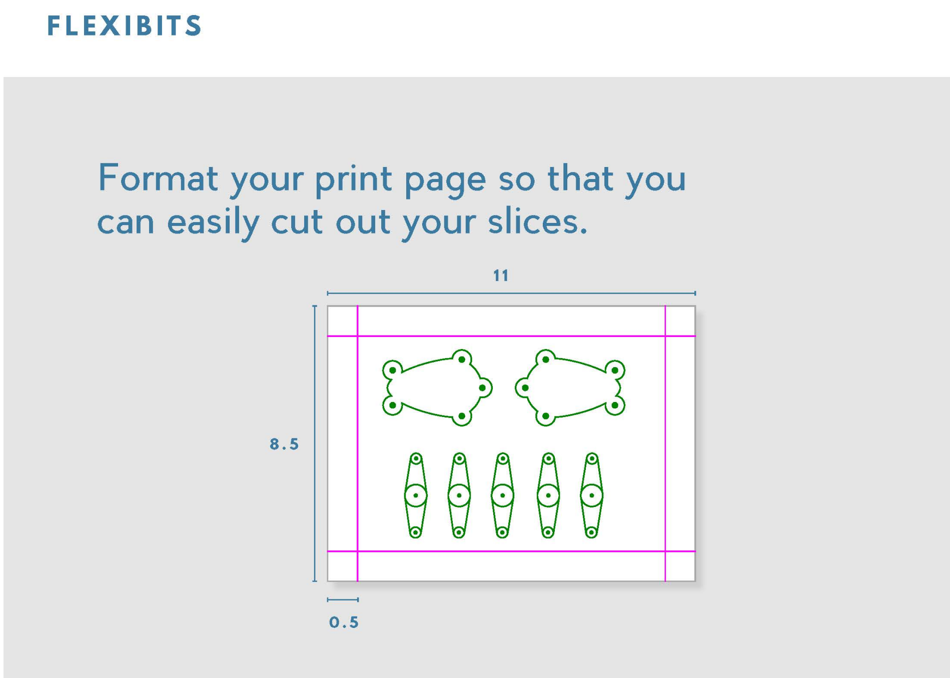


Figure A.22: FlexiBit design system explainer, page 5.

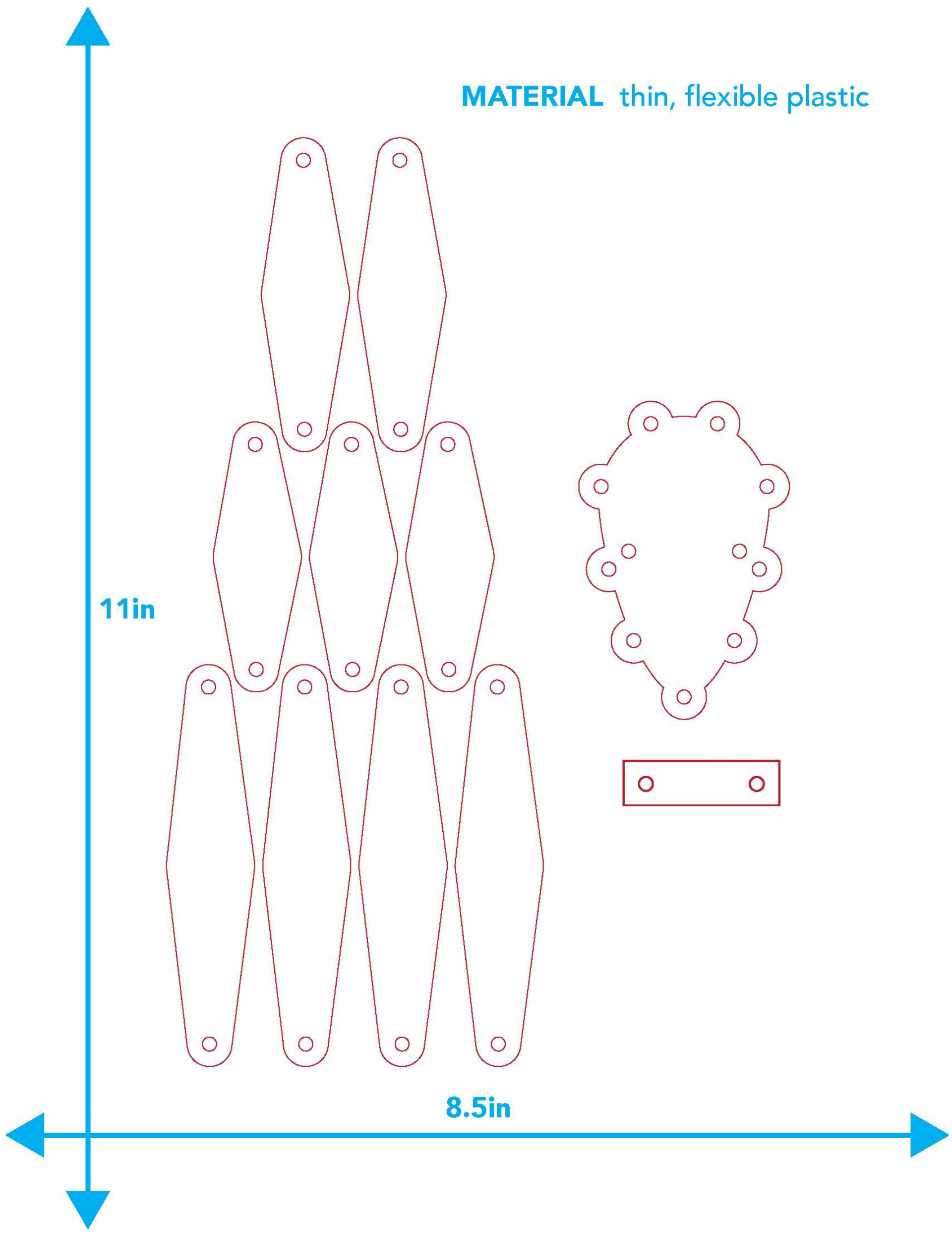


Figure A.23: FlexiBit design files to be cut out.