Voodle: vocal doodling to sketch affective robot motion

David Marino[1], **Paul Bucci**[1], **Oliver S. Schneider**[1,2], **Karon E. MacLean**[1]

[1]University of British Columbia
Vancouver, Canada
{dmarino,pbucci,maclean}@cs.ubc.ca

[2]Hasso Plattner Institute
Potsdam, Germany
oliver.schneider@hpi.de

ABSTRACT

Social robots must be believable to be effective; but creating believable, affectively expressive robot behaviours requires time and skill. Inspired by the directness with which performers use their voices to craft characters, we introduce Voodle (vocal doodling), which uses the *form* of utterances – e.g., tone and rhythm – to puppet and eventually control robot motion. Voodle offers an improvisational platform capable of conveying hard-to-express ideas like emotion. We created a working Voodle system by collecting a set of vocal features and associated robot motions, then incorporating them into a prototype for sketching robot behaviour. We explored and refined Voodle's expressive capacity by engaging expert performers in an iterative design process. We found that users develop a personal language with Voodle; that a vocalization's meaning changed with narrative context; and that voodling imparts a sense of life to the robot, inviting designers to suspend disbelief and engage in a playful, conversational style of design.

ACM Classification Keywords

H.5.1 Multimedia Information Systems: Audio Input/Output; H.5.2 User Interfaces: Natural language, User-Centered design, Voice I/O; H.5.5. Sound and Music Computing: Signal analysis, synthesis and processing

Author Keywords

Vocal interfaces; voice input; sound symbolism; animation; human-robot interaction; haptics; vocal-haptic interface

INTRODUCTION

Interactive agents are more compelling when they are believable: giving the illusion of life and facilitating suspension of disbelief [5]. 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?

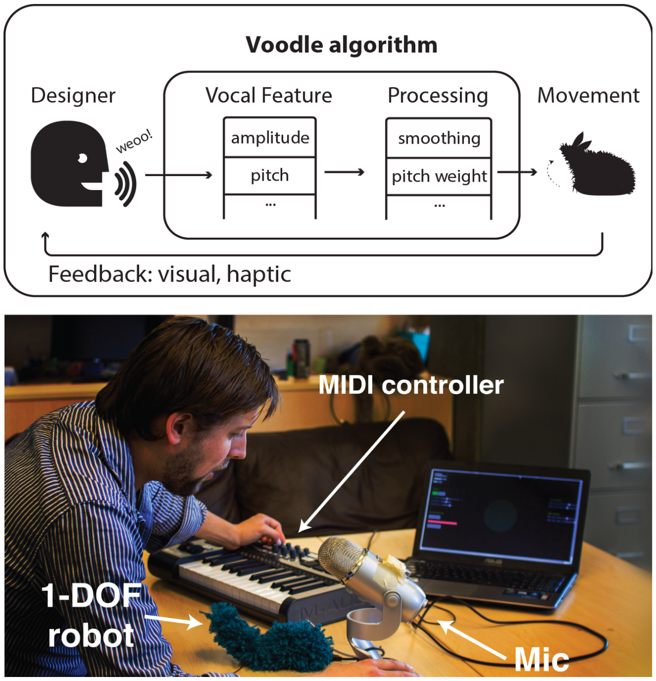


Figure 1: Voodle (vocal doodling) uses vocal performance to create believable, affective robot behaviour. Features like tone and rhythm are translated to influence a robot's movement.

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 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[1]. Specifically, we used computationally low-

DIS '17, June 10–14, 2017, Edinburgh, UK
ACM xxx-x-xxxx-xxxx-x/xx/xx. . . $15.00
DOI: http://dx.doi.org/xx.xxxx/xxxxxxx.xxxxxxx

[1] 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.

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.

Objectives, Approach and Contributions

Our goal was to support creation of believable robot behaviour through iconic 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 1 with 10 naïve users to situate Voodle in relation to traditional animation tools; and (iii) a co-design Study 2 with 3 expert performers in a 6-week intensive relationship while the Voodle system was iteratively revised.

After reviewing related work, we describe Voodle's design and implementation, then detail and reflect upon our two studies. We discuss how Voodle may fit into conventional creative processes in fields such as film, animation, theatre, or graphic design with a creative director, artist, and technician/observer; and lay out future steps, including how this experience suggests that Voodle could be extended. 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.

A *vision* for how Voodle can fit into the behaviour-design process, including when to use sound symbolic input; the domains in which sound symbolic interaction excels or is ineffective; and how voodling may integrate into a traditional design process with artist, director, and technician.

RELATED WORK

Physical Social Robots and Emotion Display

As robots enter our homes and workplaces, they need to be socially communicative. Social robots can interpret and display emotions through many modalities, including speech, touch, facial expressions, and body language [19, 62]. Physically oriented social robots, the focus of this work, have been shown to have value for companionship: Paro, a cute actuated harp seal with soft fur, has helped to manage dementia and encourage socialization in elder care homes [58]. They have potential as therapeutic tools: a fuzzy, breathing, touch-based Haptic Creature [62] helps people measurably relax [52].

These benefits come from robot movement, which must *look*—and *feel*—right. *Believability* is an essential trait for a social agent [5]; this 'illusion of life' is produced when the agent shows emotion and a thought process behind it.

Creating expressive movement – Designing expressive behaviours can be challenging, requiring animation, behavior and robot expertise as well as diverse tools [23]. 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 [15]. 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 [15]; 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 [28]; Croft and Moon mimic human hesitation behaviours on a robot [37]. Takayama applied traditional animation techniques (such as easing in/out) and tested user perceptions of robot behaviour in a video-based simulated environment [57]. Some generative techniques for affect exist as well: adding Perlin noise to robot poses can increase user recognition rate of displayed emotion [6].

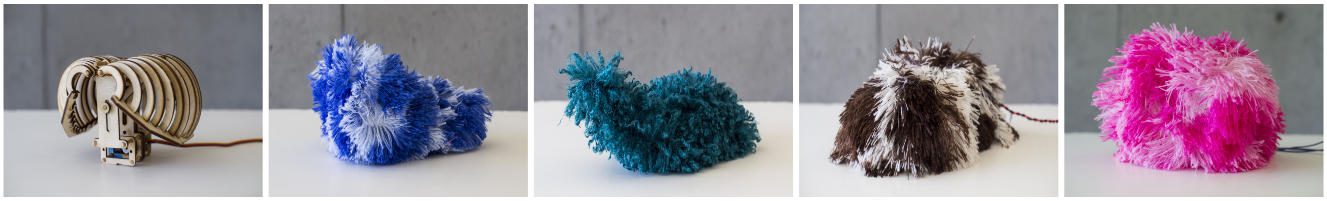
Here, we responded to individuals' natural vocal expressions by providing a novel, direct input mechanism; effectively producing *commands* that modify a pre-determined motion.

Emotion model – To study the display of emotion, we assume a conventional dimensional model for emotion, based on Russell's circumplex model for affect [46], which places emotions on a two-dimensional plane. Valence (pleasant/unpleasant) is along one axis, and arousal (high/low activation) is orthogonal. Russell's circumplex model is often discretized into a grid [45], or represented by a set of validated words such as the Positive and Negative Affect Schedule (PANAS, [61]). In PANAS, words are grouped by various configurations of the affect grid, such as by quadrant, and assumed to be roughly equivalent in affective intensity. For example, "stressed" is (high-arousal, low-valence) along with "upset" or "angry" and is opposed diagonally by "relaxed" (low-arousal, high-valence) or "serene". The circumplex model is useful in describing dynamic emotional transitions; transitions were easier for participants to perform over single emotions in our study.

Expressive capacity of simple agents – While affective robots have successfully expressed complex emotional behaviours on human-like platforms (e.g., facial expressions [9]), humans also have a powerful ability to anthropomorphize, easily constructing narratives and ascribing complex emotions to non-human objects [26] — even a simple rod in motion [25].

We have leveraged this ability with 1-degree-of-freedom (DOF) robot breathing behaviours that can recognizably render diverse emotions [12, 62], most recently in our simple but expressive zoomorphic CuddleBit robot family [12]. In a previous study, we found that CuddleBits could consistently express a variety of emotions across the affect grid [12].

However, the design space of even these 1-DOF robots is too large to traverse with conventional tools like keyframe editors. We had anecdotally observed individuals struggle with more conventional behavior-generation approaches, and



(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.

Figure 2: The 1-DOF CuddleBit robots used in co-design Study 2.

instinctively turn to iconic vocalizations to describe what they wanted. The idea of Voodle came from the sense that it should be possible to use those vocalizations more directly.

Voice Interaction

Alignment – A fundamental component of natural human communication is *alignment* between conversation partners, which occurs when people mimic one another's communicative patterns [20]. Phonetic *convergence* refers specifically to alignment of speakers' phonetic patterning [38], and other studies show that people similarly coordinate speech rhythm, body language and breathing pattern [33,36,59]. Similarly, mimicry positively impacts affiliation and likability [33]. Alignment extends to human-computer conversations: people adjust language to their expectations of how a system works [39].

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 [3,22,35]; Hoffman has explored human-robot alignment by utilizing computer vision techniques within performative contexts [27]. Here, we use iconic features of the speech signal to achieve the illusion of alignment, making for more believable interaction.

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) [18]. 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 [40–42]: 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”*) [41].

Iconic vocalizations carry emotional content. Banse and Scherer found that iconic voicing excels in communicating psychological phenomena such as emotional states [4]; 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) [44].

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 [49,60], and to ground and communicate design intention [2,11]. 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 [54]. 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 [21,31], enhance accessibility of computer systems [8], or as intuitive input for artistic expression [16,30]. Voice Augmented Manipulation augments users' touch input with voice, e.g., as a modal modifier key [48]. 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 [10].

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.

SYSTEM DESIGN

To explore voodling, we built a design input tool that translates voodles to socially expressive motion for the CuddleBit family of 1-DOF robots (Figure 2) [13]. 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.

Pilot Study: Gathering Requirements

We conducted a pilot to inform an initial Voodle implementation based on the RibBit (Figure 2). Like most of the Bits, the RibBit moves its “ribs” in and out with a breathing-like motion.

Table 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.

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.

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 [53]. 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 [51].

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

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.

Results

Phonetic Features – Table 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 [14]. 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.

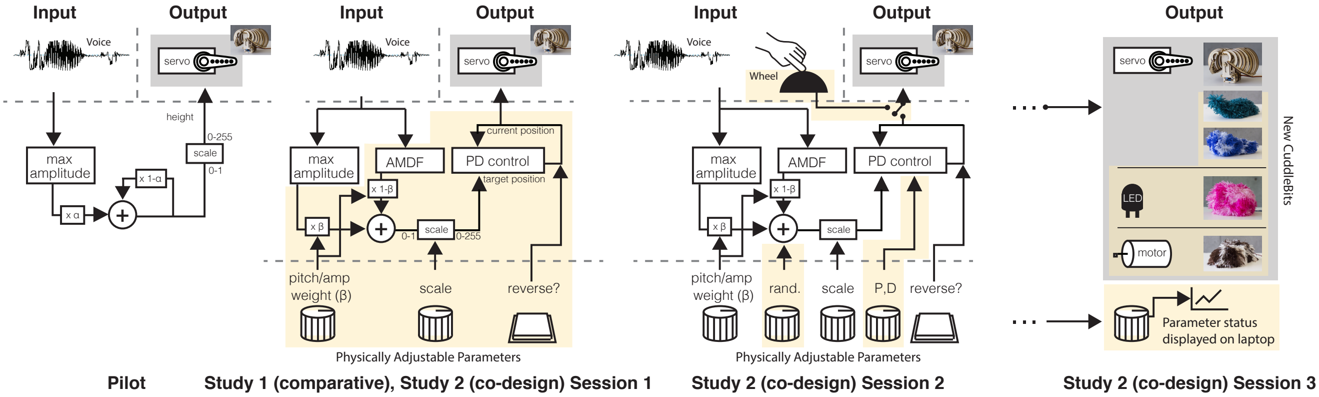
There were exceptions to this pattern; for example, one participant saw the robot's downward movement as 'flexing,' and therefore used increased vocal pitch and amplitude to represent its downward movement. Table 1 reports contrasting relationships that we observed, with examples.

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"*).

Voodle Implementation

Based on piloting guidance, we created a full Voodle system, seen in Figures 3 (system design) and 1 (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)



Pilot Study 1 (comparative), Study 2 (co-design) Session 1 Study 2 (co-design) Session 2 Study 2 (co-design) Session 3

Figure 3: 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.

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 1.

Voodle was implemented in JavaScript: a NodeJS server connected with the RibBit using Johnny-Five and ReactJS [32,43]. 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 [55], the best performer in informal piloting. Figure 3 shows algorithm evolution. Voodle is open-source, available at https://github.com/ubcspin/Voodle.

STUDIES

To understand and develop the possibilities of voodling within a creative design process, we conducted two studies.

In the first, we examined Voodle's role as a design tool by comparing it to traditional digital animation methods, such as keyframing. To do this, we situated Voodle within established performative design practices, as seen in voice acting, puppeting, and animation. In these disciplines it is common to employ a director-artist-technician setup to ensure quality: the director delivers detached artistic feedback to the artist, while the technician can offer technical advice and suggestions.

The second activity was a co-design process: 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. For both studies, methods borrowed from grounded theory [56] 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.

COMPARATIVE STUDY 1: Situating Voodle as a Tool

Procedure

We recruited ten participants to design five robot behaviours, each based on an emotion word from the PANAS scale [61]. 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. 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 [50], 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.

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, 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 [12]. 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. To understand the voodling experience and improve its implementation, we conducted a co-design activity (Study 2) with expert performers who could push its expressive capability.

CO-DESIGN STUDY 2: Performer Use and Revision

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.

Methods – 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* → *relaxed*, for (high-arousal, high-valence) → (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 4). (3) pile-sort [7] 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.

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.

Analysis – We conducted thematic analysis [47] informed by grounded theory methods [17] on observations, video, and interview data. We found four themes (Table 2): 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.

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 (∼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* → *depressed*, P2 started by saying *"I'm having a nice relaxing*

Table 2: Summary of Study 2 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

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. ⌈**KM** *SLC*⌋

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* [34].

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.

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.

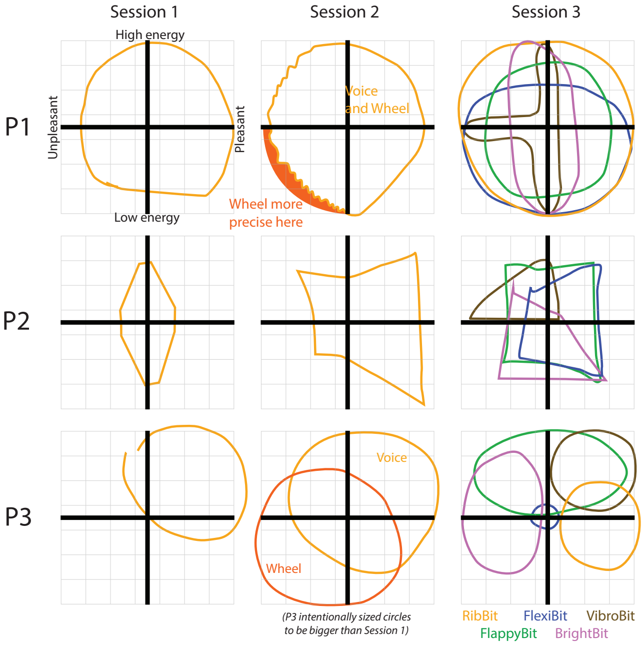
After some experimentation, P1 incorporated more pitch control, which afforded better control. P2 and P3 indicated increased expressivity on their affect grids in Session 2 (Figure 4), suggesting improvement of either ability or system.

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"* (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 4).



(P3 intentionally sized circles to be bigger than Session 1)

Figure 4: 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.

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 2).

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 2), 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 4), 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 4). 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.

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.

DISCUSSION

Our initial goal was to create a dedicated design tool for affective robots. From these studies, 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.

Insights into Believability and Interactivity

Through co-design Study 2, 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 [20]; 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.

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 [1, 23]). 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 co-design Study 2, 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 [12, 29], 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 [8, 24]), which had a defined, learned control space. As we discovered in our pilot and confirmed in comparative Study 1, 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 Voodle

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 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 (comparative Study 1), 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 comparative Study 1 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.

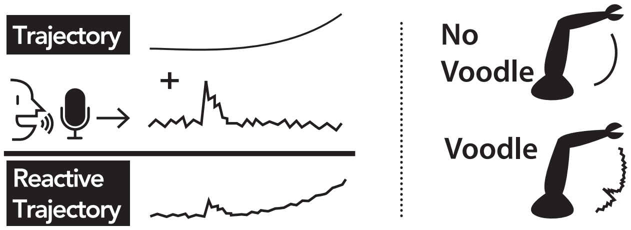


Figure 5: Vision for "Embedded Voodle": Voodle could be a natural low-cost method to emotionally color a motion path in more complex robots.

Iconic vocalizations have also been used to describe tactile sensations [49, 60], 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 [6], 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). 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*).

CONCLUSION

In this paper, we introduced Voodle: vocal doodling, an interaction technique using iconic input. We collected a set of iconic vocalizations from users and linked them to robot behaviours. This informed our implemented Voodle system, which maps pitch and amplitude to robot motion in an extensible, parameterized algorithm. In two studies, we 1) found that Voodle is not a stand-alone design tool, but can be combined with a keyframe editor to create expressive robot behaviours, and 2) identified themes of personal iconic language, narrative frame, alignment, and control methods. We found that voodling is a blend of social robot interaction and puppetry-based design, with the potential to add a "spark of life" to physical interactive systems.

ACKNOWLEDGMENTS

We thank our study participants and reviewers. This work was supported by NSERC, and performed under UBC Ethics H13-01620.

REFERENCES

1. Anki. 2017. (2017). https://anki.com/en-us/cozmo.

2. Farah Arab, Sabrina Paneels, Margarita Anastassova, Stephanie Coeugnet, Fanny Le Morellec, Aurelie Dommes, and Aline Chevalier. 2015. Haptic patterns and older adults: To repeat or not to repeat?. In *2015 IEEE World Haptics Conference (WHC)*. IEEE, 248–253. DOI: http://dx.doi.org/10.1109/WHC.2015.7177721

3. Jeremy N Bailenson and Nick Yee. 2005. Digital chameleons automatic assimilation of nonverbal gestures in immersive virtual environments. *Psychological science* 16, 10 (2005), 814–819.

4. Rainer Banse and Klaus R Scherer. 1996. Acoustic profiles in vocal emotion expression. *Journal of personality and social psychology* 70, 3 (1996), 614.

5. Joseph Bates and others. 1994. The role of emotion in believable agents. *Commun. ACM* 37, 7 (1994), 122–125.

6. Aryel Beck, Antoine Hiolle, and Lola Canamero. 2013. Using perlin noise to generate emotional expressions in a robot. In *Proceedings of annual meeting of the cognitive science society (Cog Sci 2013)*. 1845–1850.

7. H Russell Bernard. 2011. *Research methods in anthropology: Qualitative and quantitative approaches*. Rowman Altamira.

8. Jeff A Bilmes, Xiao Li, Jonathan Malkin, Kelley Kilanski, Richard Wright, Katrin Kirchhoff, Amarnag Subramanya, Susumu Harada, James A Landay, Patricia Dowden, and Howard Chizeck. 2005. 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)*. Association for Computational Linguistics, Stroudsburg, PA, USA, 995–1002. DOI: http://dx.doi.org/10.3115/1220575.1220700

9. Cynthia Breazeal. 2003. Emotion and sociable humanoid robots. *International Journal of Human-Computer Studies* 59, 1-2 (jul 2003), 119–155. http://dx.doi.org/10.1016/S1071-5819(03)00018-1

10. Cynthia Breazeal and Lijin Aryananda. 2002. Recognition of Affective Communicative Intent in Robot-Directed Speech. *Autonomous Robots* 12, 1 (2002), 83–104. DOI: http://dx.doi.org/10.1023/A:1013215010749

11. L. Brunet, C. Megard, S. Paneels, G. Changeon, J. Lozada, M. P. Daniel, and F. Darses. 2013. "Invitation to the voyage": The design of tactile metaphors to fulfill occasional travelers' needs in transportation networks. In *2013 World Haptics Conference (WHC)*. IEEE, 259–264. DOI:http://dx.doi.org/10.1109/WHC.2013.6548418

12. Paul Bucci, Laura Cang, Sazi Valair, David Marino, Lucia Tseng, Merel Jung, Jussi Rantala, Oliver Schneider, and Karon MacLean. 2017. 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).

13. Paul Bucci, Xi Laura Cang, Matthew Chun, David Marino, Oliver Schneider, Hasti Seifi, and Karon MacLean. 2016. CuddleBits: an iterative prototyping platform for complex haptic display. In *EuroHaptics '16 Demos*.

14. Noam Chomsky and Morris Halle. 1968. *The sound pattern of English.* New York: Harper and Row. 435 pages.

15. Howie Choset, Kevin M Lynch, Seth Hutchinson, G Kantor, Wolfram Burgard, Lydia E Kavraki, and Sebastian Thrun. 2005. *Principles of Robot Motion: Theory, Algorithms, and Implementations*. The MIT Press.

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

17. Juliet Corbin and Anselm Strauss. 2008. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory* (3 ed.). Sage Publications, Inc. 379 pages.

18. Ferdinand De Saussure, Wade Baskin, and Perry Meisel. 2011. *Course in general linguistics*. Columbia University Press. 336 pages.

19. Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. 2003. A survey of socially interactive robots. *Robotics and Autonomous Systems* 42, 3 - 4 (2003), 143–166. DOI: http://dx.doi.org/10.1016/S0921-8890(02)00372-X

20. Simon Garrod and Martin J Pickering. 2004. Why is conversation so easy? *Trends in cognitive sciences* 8, 1 (2004), 8–11.

21. Masataka Goto, Koji Kitayama, Katunobu Itou, and Tetsunori Kobayashi. 2004. Speech Spotter: On-demand Speech Recognition. In *in Human-Human Conversation on the Telephone or in Face-to-Face Situations. Proc. ICSLP'04*. 1533–1536.

22. Jonathan Gratch, Anna Okhmatovskaia, Francois Lamothe, Stacy Marsella, Mathieu Morales, Rick J van der Werf, and Louis-Philippe Morency. 2006. Virtual rapport. In *International Workshop on Intelligent Virtual Agents*. Springer, 14–27.

23. Jesse Gray, Guy Hoffman, Sigurdur Orn Adalgeirsson, Matt Berlin, and Cynthia Breazeal. 2010. 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*.

24. Susumu Harada, James A Landay, Jonathan Malkin, Xiao Li, and Jeff A Bilmes. 2006. 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)*. ACM, New York, NY, USA, 197–204. DOI: http://dx.doi.org/10.1145/1168987.1169021

25. John Harris and Ehud Sharlin. 2011. Exploring the affect of abstract motion in social human-robot interaction. In *2011 Ro-Man*. IEEE, 441–448.

26. Fritz Heider and Marianne Simmel. 1944. An experimental study of apparent behavior. *The American Journal of Psychology* 57, 2 (1944), 243–259.

27. Guy Hoffman. 2011. 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*, Vol. 1.

28. Guy Hoffman and Wendy Ju. 2014a. Designing robots with movement in mind. *Journal of Human-Robot Interaction* 3, 1 (2014), 89–122.

29. Guy Hoffman and Wendy Ju. 2014b. Designing robots with movement in mind. *Journal of Human-Robot Interaction* 3, 1 (2014), 89–122.

30. Brandi House, Jonathan Malkin, and Jeff Bilmes. 2009. The VoiceBot: A Voice Controlled Robot Arm. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '09)*. ACM, New York, NY, USA, 183–192. DOI: http://dx.doi.org/10.1145/1518701.1518731

31. Takeo Igarashi and John F Hughes. 2001. 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)*. ACM, New York, NY, USA, 155–156. DOI: http://dx.doi.org/10.1145/502348.502372

32. Johnny-Five. 2017. (2017). http://johnny-five.io.

33. Jessica L Lakin, Valerie E Jefferis, Clara Michelle Cheng, and Tanya L Chartrand. 2003. The chameleon effect as social glue: Evidence for the evolutionary significance of nonconscious mimicry. *Journal of nonverbal behavior* 27, 3 (2003), 145–162.

34. George Lakoff and Mark Johnson. 2003. *Metaphors we live by*. 276 pages.

35. RM Maatman, Jonathan Gratch, and Stacy Marsella. 2005. Natural behavior of a listening agent. In *International Workshop on Intelligent Virtual Agents*. Springer, 25–36.

36. David H McFarland. 2001. Respiratory markers of conversational interaction. *Journal of Speech, Language, and Hearing Research* 44, 1 (2001), 128–143.

37. Ajung Moon, Chris AC Parker, Elizabeth A Croft, and HF Machiel Van der Loos. 2011. 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*. IEEE, 1994–1999.

38. Jennifer S Pardo. 2006. On phonetic convergence during conversational interaction. *The Journal of the Acoustical Society of America* 119, 4 (2006), 2382–2393.

39. Jamie Pearson, Jiang Hu, Holly P Branigan, Martin J Pickering, and Clifford I Nass. 2006. 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*. ACM, 1177–1180.

40. CS Peirce. 1955. *The philosophical writings of Peirce*. New York: Dover, 98–119.

41. Marcus Perlman and Ashley A Cain. 2014. Iconicity in vocalization, comparisons with gesture, and implications for theories on the evolution of language. *Gesture* 14, 3 (2014), 320–350.

42. Pamela Perniss and Gabriella Vigliocco. 2014. The bridge of iconicity: from a world of experience to the experience of language. *Phil. Trans. R. Soc. B* 369, 1651 (2014), 20130300.

43. React. 2017. (2017). https://facebook.github.io/react.

44. Ralf Rummer, Judith Schweppe, René Schlegelmilch, and Martine Grice. 2014. Mood is linked to vowel type: The role of articulatory movements. *Emotion* 14, 2 (2014), 246.

45. James A Russel, Anna Weiss, and Gerald A Mendelsohn. 1989. Affect grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology* 57, 3 (1989), 493–502.

46. James A Russell. 1980. A circumplex model of affect. *Journal of Personality and Social Psychology* 39, 6 (1980), 1161.

47. Gery W. Ryan and H. Russell Bernard. 2003. Techniques to Identify Themes. *Field Methods* 15, 1 (feb 2003), 85–109. DOI: http://dx.doi.org/10.1177/1525822X02239569

48. Daisuke Sakamoto, Takanori Komatsu, and Takeo Igarashi. 2013. 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)*. ACM, New York, NY, USA, 69–78. DOI:http://dx.doi.org/10.1145/2493190.2493244

49. Oliver S Schneider and Karon E MacLean. 2014. Improvising design with a haptic instrument. In *2014 IEEE Haptics Symposium (HAPTICS)*. IEEE, 327–332.

50. Oliver S. Schneider and Karon E. MacLean. 2016. 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*.

51. Oliver S. Schneider, Hasti Seifi, Salma Kashani, Matthew Chun, and Karon E. MacLean. 2016. HapTurk: Crowdsourcing Affective Ratings for Vibrotactile Icons. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI) '16*. ACM Press, New York, New York, USA, 3248–3260. DOI: http://dx.doi.org/10.1145/2858036.2858279

52. Yasaman S Sefidgar, Karon E MacLean, Steve Yohanan, HF Machiel Van der Loos, Elizabeth A Croft, and E Jane Garland. 2016. Design and Evaluation of a Touch-Centered Calming Interaction with a Social Robot. *IEEE Transactions on Affective Computing* 7, 2 (2016), 108–121.

53. Hasti Seifi, Kailun Zhang, and Karon E MacLean. 2015. VibViz: Organizing, visualizing and navigating vibration libraries. In *World Haptics Conference (WHC), 2015 IEEE*. IEEE, 254–259.

54. Hadas Shintel, Howard C Nusbaum, and Arika Okrent. 2006. Analog acoustic expression in speech communication. *Journal of Memory and Language* 55, 2 (2006), 167–177.

55. Joren Six, Olmo Cornelis, and Marc Leman. 2014. TarsosDSP, a Real-Time Audio Processing Framework in Java. In *Proceedings of the 53rd AES Conference (AES 53rd)*.

56. Anselm Strauss and Juliet Corbin. 1998. *Basics of qualitative research: Techniques and procedures for developing grounded theory* . Sage Publications, Inc.

57. Leila Takayama, Doug Dooley, and Wendy Ju. 2011. Expressing thought: improving robot readability with animation principles. In *Proceedings of the 6th international conference on Human-robot interaction*. ACM, 69–76.

58. Kazuyoshi Wada and Takanori Shibata. 2007. Living with seal robotsâĂŤits sociopsychological and physiological influences on the elderly at a care house. *IEEE Transactions on Robotics* 23, 5 (2007), 972–980.

59. Rebecca M Warner. 1996. Coordinated cycles in behavior and physiology during face-to-face social interactions. (1996).

60. Junji Watanabe and Maki Sakamoto. 2012. 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)*. 2346–2348.

61. David Watson, Lee A Clark, and Auke Tellegen. 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of personality and social psychology* 54, 6 (1988), 1063.

62. Steve Yohanan and Karon E MacLean. 2011. Design and assessment of the haptic creature's affect display. In *Proceedings of the 6th international conference on Human-robot interaction*. ACM, 473–480.