www.icgst.com

Emotionally Expressive Music based Interaction Language for Social Robots

Aladdin Ayesh
School of Computer Technology, De Montfort University, The Gateway, Leicester LE1 9BH
email:aayesh@dmu.ac.uk,
`http://www.aladdin-ayesh.info`

Abstract

Sound is perhaps the most elementary and yet common communication vehicle used by humans and animals alike. Similarly, reactive robots that are based on animal intelligence require real-time simple communication mechanisms to imitate their animal examples. They would require their own natural language in which sound blocks can be contained in meaningful structures without the complexity of human's natural language. Furthermore, this natural language of animal like robots should be scalable to evolve depending on the available resources to each species of robots. In this paper, a musical language is proposed to fulfil this requirement. Specifications of the language and case studies of using these specifications are provided to demonstrate its applicability in multi-robots and human-robot interactions.

Keywords: Emergent behavior, emotional interaction, animal intelligence, auditory interaction.

1 Introduction

Behavioral, verbal and non-verbal expressions for interacting are relatively easier in deterministic systems than in non-deterministic systems. Non-deterministic systems, however, have increasing uses in variety of engineering and domestic applications. At the same time, many of these systems, often in the form of robots, have limited resources to handle both action selection and natural language communication, let alone representing underlying subtleties of interacting such as emotions and uncertainties. Expressing emotions in particular has many applications in education and domestic robots [46, 7, 40]. Emotion modeling is often associated with cognitive agents that are capable of reasoning about themselves [15]. One may argue, however, that emotions are reactive triggers and thus suited for reactive agents [21, 14]. It can be argued that emotions provide the means of non-deterministic interaction emulating animals and humans. Furthermore, emotions models can vary in complexity to reflect the complexity of psychological and sociological internal representations. Taking this viewpoint this paper presents the development of a musical language for communication and emotional expression that can be used in reactive robots and extended to cognitive agents. The idea is to give robots their own *natural* language for expression in imitation to the animals' behavior in their use of sounds to express themselves. Using a musical language allows us to draw on a large literature from musicology and musical technology [3, 43, 39, 50, 51, 27, 2]. In fact, [39] provides an interesting sound propagation model for communication between agents. Such work may form a sound manipulation base for our work that is presented here whilst our work provides a high level representational language. The paper does not aim, however, to discuss or to develop the sound recognition and synthesis at this stage. Instead, our aim is to present operators for the synthesis and analysis of musical messages to provide an alternative language to speech acts with scalable formats and emotionally expressive. The use of musical language provides musical syntax and grammar to enable the development of an interactive and emotionally expressive language whereby the complexity of the grammar can be easily maintained at a simple level that suits reactive robots. The main part of this paper discusses the syntax of our proposed Musical Language for Emotional Interaction between Robots ($MLEIR$). The rest of the paper is organized as follows: it starts with a background in section 2 establishing the motivations and applications of the proposed language. Sections 3.1 and 4 provide the syntax of the musical language (MLEIR) and its synthetic operators. Both extended BNF and Z notations are used in developing MLEIR. Section 5 provides the workings of the language in relation to mental and emotional states and shows the use of MLEIR through case study that is elaborated further in section 7. This is followed by analytical analysis providing a preliminary assessment of MLEIR.

2 Motivations

2.1 Human Computer Interaction (HCI)

The benefits of using sound in general and music in particular as communications mechanism in robots may be questionable. First question may come to mind, what is the difference between using musical sound and normal signaling? The answer is that in signaling (including sound signaling) straightforward messages that are deterministic and exact are encoded. Living creatures are rarely do so except perhaps in the socially disciplined species such as ants. Even humans with their highly developed natural languages the underlying tonal expression in the human voice delivers more than the spoken message. However, human's vocal expressions are far more complex and vary, it may prove difficult to analyze and to model. Especially if one considered influences of the facial expressions [4, 54, 55] and hand-body gestures on message interpretation [23, 24, 25, 28, 29, 30, 34, 41]. Another question is that what is the added value of using musical language? There are several benefits of using musical language that may become more apparent If one considers the practical applications of this research in providing some alternative mechanisms in human-machine interaction. Music has the advantage of combining the simple sound, which can be used in basic signal communication, and the structure of a language that can be extended and modified. From a human viewpoint, musical tones and sentences are often easy to recognize and remember [46]. Sound is already used in many toys, especially with focus on children developmental learning. There is a particular interest in using sound, rather than synthesis voices, when it comes to autistic children [44, 45]. In addition, there is a continued development of computer software and research on the use of music for communicating information [51, 2]. As for musical composition, active and dynamic music composition methods still strongly fascinate both scientists and musicians [18, 38, 53, 8]

What is proposed here is to use sound in a structured way to develop a synthetic abstract language of communication that can be substituted depending on the application. Music by its nature enables abstraction and substitution. Hence, the use of 'music' here to describe the language to indicate both auditory sound, notational representation, and musical information processes [8], [11], [33], [37], [35]. The conceptual idea has been recognised in literature [51], [2] and [32] but this paper attempts to formalise this concept in a generic framework beyond what proposed in literature, e.g. emotional content of musical phrases as mentioned in [32]. The importance of the emotional content of communicative messages is well recognised in psychological and human-robot literature that commands a survey by itself, nonetheless few but to mention examples are [10, 13, 48, 49].

2.2 Animal Intelligence based Multi-Robots

Communication between animals takes various forms. There is *smell* that is used during the mating period, to mark and identify territories, to inform oneself and other animals of the route taken and so on. There is also *touch*, which is the easiest to model. Finally, there is *sound*, which is commonly used with a variety of frequencies [31] some of which are inaudible to the human ear. Observing animals that use sound for communication such as dogs and cats one can identify certain relationships between the sound produced and the feelings of that animal. The uncomfortable dog may start howling, whilst stepping on his toe may lead to a sharp short bark expressing urgency. Similarly, reactive robots that model animal intelligence [9, 14, 52] should be able to demonstrate the same capabilities of own sound language for expression to enable realistic simulations and interactions. In other words, a 'natural language like' of little animal like living robots. Esnaola and Smithers [20] seem to agree with the notion presented here. The main difference is that they design their musical language based on communication packets used with cross fertilization between multi-agent systems communication protocols and natural language. Whilst the results of great interest, one can see very little organic expansion possible. In addition, there was no clear indication of emotions representation or expression of emotional content. One may argue that their musical language is a good replacement to speech acts of multi-agent systems.

The main aim is to allow the expression of underlying information as in natural languages where emotions and other expressive information is encoded implicitly to enable natural interaction between robots and humans in realization of social robots [19]. Thus, emotions will be encoded in the communicative messages rather than expressed as part of the message structure. This will enable the simplification of message structure. Music is very suited for such endeavor for many reasons. Firstly, music is formed of sounds which allows us to imitate the natural animal sound-based communication. Secondly, music is associated with emotions [46] and therefore it is easy to encode emotions within musical messages at no extra complexity and with great applicability to human-computer interaction. Finally, systematic analysis of music and musical languages [12, 17]have been studied and grammar-based musical representation has been developed [37]. This gives us the confidence in the possibility of expanding our musical language.

3 Musical Language Design

Sound is often modeled as a signal for which neural networks and sound processing techniques are usually used, e.g. [16], [35], [36], and [56], to mention but few. However, let us look at sound as structured phonetics, these structures are associated in natural languages with characters and words. The closest example is pictorial languages where each image has a sound and a meaning. A slight variation on the sound changes the meaning. In designing the language here, our interest is to define the structures in which the sound properties can be accommodated, giving them a meaning, and a syntactical (and eventually semantical) association with other structures. Developing such a language is a huge task and so the following limitations applied: (1) Use of Western tonal music theory as the basis for syntax and grammar; (2) Use of formal method representation; (3) Limiting discussion on semantics. The following subsection provides the syntactical description and grammar rules used whilst the synthetic operators will be covered in a separate section.

3.1 MLEIR Syntactic Specifications

The notation used here is based on Z notation [42] and Extended BNF [26]. There are four types of terminal symbols defined, which are sound class (SC), sound pitch (SP), sound duration (SD), and sound volume (SV).

$$SC : \{'C'|'D'|'E'|'F'|'G'|'A'|'B'\}$$

Sound class refers to the basic musical sounds that can be found on the white keys of a piano. The term sound classes is to highlight the fact that all other musical sounds are driven from these classes by varying their features such as pitch or register. This includes adding flats and sharps to generate new sounds, which can be treated as new classes. For practical reasons, dealing with flats and sharps is omitted from this paper.

$$SP : \{'1'|'2'|'3'|'4'|'5'|'6'|'7'\}$$

Pitch numbers are effectively a register reference whereby C3 is perceived an octave lower than C4.

$$SD : \{'1'|'2'|'4'|'8'|'16'\}$$

Where 1, 2, 4, 8 and 16 refer to semibreve, minim, crotchet, quaver, and semiquaver, respectively.

$$SV : \{'1'|'2'|'3'\}$$

Whereby 1, 2 and 3 denotes ranges of loudness related to emotional expressions with 1 being quiet range, which may be expressing fear for example, 2 being natural or neutral and 3 being loud, which may be expressing anger.

If one wishes to apply Z type system, taking sound S to be the main type set one could write:

$$S = \{sc\ is\ sound_{sp} \mid sc, sp \equiv_{def} \langle SD \times SV \rangle\} \quad (1)$$

The addition of *sp* to sound in the formula 1 is to indicate that *sp* is strongly associated with *sc*, whilst *sd* and *sv* defines the sound as features.

A sound message (SM) can then be constructed from these terminal symbols in the following format:

$$SM = \langle SC, SP \rangle \quad (2)$$

$$SM = \langle SC, SP, SD \rangle \quad (3)$$

$$SM = \langle SC, SP, SD, SV \rangle \quad (4)$$

Formula 2 provides the naming of the robot emitting the message. Section 3.3 discusses naming. These formulas may be summarized as follows:

$$\forall s : S \cdot s \equiv_{def} \{\langle SC \times SP \rangle \times [SD] \times [SV]\} \quad (5)$$

3.2 Conversation

A conversation is constructed through a sequence of *SM*. Two definitions of conversation are provided here. Definition 1 describes the conversation sequence, which is the basis of communication whilst definition 2 establishes conversational limits and eliminates conversational conflicts.

Definition 1 (Conversation Sequence) *A conversation may be defined to be an array of SM* $\{SM_1 \ldots SM_i\}$ *whereby* $SC_1 = SC_2 = \ldots = SC_i$

Definition 2 (Conversation Constraint) *Formally we define conversation* Ψ *with perception constraint as follows:*

$$\exists \alpha \cdot \exists \beta : \Gamma \cdot \Psi \subseteq \{*SM_{\alpha \mapsto \beta} \times *SM_{\beta \mapsto \alpha}\}_{t_0 \succ t_n} \Longrightarrow \exists \chi : \Gamma \cdot SM_{\chi \mapsto \alpha \vee \beta} \longleftarrow \{\}_{t_0 \succ t_n} \quad (6)$$

Where Γ is the group of robots (agents), $t_0 \succ t_n$ indicates the temporal limits of the conversation Ψ between robot α and robot β whilst $\longleftarrow \{\}_{t_0 \succ t_n}$ denotes a no response mode to robot χ communication attempt.

Both definitions are used in delineating synthetic communication operator in section 4.3

3.3 Naming convention

For conversation to take place a naming convention is needed. SC is used in association with SP to develop such a naming convention. SP, as we will see later on in 5, has important role to play in emotional expression. Consequently, a particular pitch is associated to be *a natural voice* of the robot. This allows the robot to use other pitches to express its emotions.

The first possible naming convention is to name each robot individually, which will be referred to as *Individual Naming Convention* (INC). Expressing the robot name in a sound message is formulated in 7. Sound class and pitch are used to identify the robot. Each robot will have a natural voice or pitch.

$$SM_{\sigma(\alpha)} = \left\langle SC_{(\alpha)}, \theta SP_{(\alpha, normal)} \right\rangle \quad (7)$$

Where σ is the naming operator and θ is the natural pitch associated with the robot being named. The idea behind natural pitch of a robot is to provide naming uniqueness, which is defined in 3

Definition 3 (Naming Uniqueness) *We define naming uniqueness to be that:*

$$\exists \alpha \cdot \exists \beta : \Gamma \cdot \sigma(\alpha) = \sigma(\beta)\ iff\ \alpha = \beta. \quad (8)$$

Alternative naming convention may take into consideration team formation or *Team Naming Convention*(TNC). In this case, each robot will be named in association to the team it follows. Each team use the formula 7 to specify team name and to specify a robot identifier. Then the two names are concatenated to produce the individual robot's name.

4 Operators

For the musical language to adaptable for different applications, generic synthetic operators are required.

4.1 Initiation: Self Operator (SfO)

Representing self is of great importance for orientation and reasoning especially when social and interaction rules apply such is the case in multi-agent systems and robot collectives. Self and its representation had a great interest from psychologists, philosophers and artificial intelligence researchers alike [5]. To avoid argumentative cases and paradoxes associated with representing self, self is represented here as a root node from which relations emerge. As a root node, it acts as a start and termination node of information as a given agent is concerned for whom this node represents self. Thus, consciousness is defined here as the maintenance of the root node of self.

4.2 Belief and Emotion Synthetic Operator (BESO)

Belief and Emotions Synthetic Operator maps the mental state of the robot to the phonetic equivalence. The intersection between Beliefs and Emotions provides the Perception and Intention. The Belief and Emotion Synthetic Operator (BESO) maps between the belief set, emotional state and the message construction which is provided by the Communication Synthetic Operator (CSO). The impact of BESO is seen in the alteration applied to properties of sound structures that are forming the musical message.

4.3 Communication Synthetic Operator (CSO)

Communication is done through the construction of a sequence of musical sounds. This sequence is effectively a conversation with a short time span allowed. The sequence is synthesized as continuous SM in response to environment stimuli. This operator has two modes: constructive mode $(CSO \rightarrow)$ for message production and interpretive mode $(CSO \leftarrow)$ for message interpretation. The one that is used in our case studies is the constructive mode since our concern is the production of sound structures rather than sound analysis as a signal. $(CSO \rightarrow)$ can be defined as follows:

$$\exists \alpha : \Gamma \cdot \exists s_i : S \cdot SM_i \mapsto s_{i+1} = CSO \rightarrow \left\{ SC_{(name_\alpha)}, Map\,(s_i) \right\} \quad (9)$$

Formula 9 specifies how the SM_i in response to the current state s_i will be formulated in the next state s_{i+1}. Note that the first part of the robot name, namely SC, is extracted and used as a message tag. In other words, a 'C' type robot will always use the note 'C' in forming its messages. This is the equivalence of dog barking, cat meowing and bird singing.

4.4 Action Selection Process (Inference Engine)

Action selection process is based on the inference mechanism used. There is no hard rule of action selection process specification. It should, however, encompass the language specifications and synthetic operators, and to be applicable to the re-writing rules (section 5). Two possible approaches to achieve this are:

Rule based inference mechanisms (e.g. forward and backward chaining)

Semantic mapping (e.g. cognitive map and graph planning)

Definition 4 gives us an abstract behavior expression rule that provides guidelines specification to the inference engine.

Definition 4 (Behavior Expression Rule)

$$\exists s_n, s_{n+1} \in S. \exists a \in A. s_n < SfO_{\Re}, BESO, CSO \longleftrightarrow > \vdash s_{n+1} < a, CSO \longrightarrow > \quad (10)$$

Conjecture 1 *Since the sound expression is in itself a behavioral utterance and similarly behaviors can be viewed as muted physical utterances of communication, the musical language can be used for the general purpose of internal representation.*

5 Emotional Expression: Re-writing Rules

One aspect of the language that is not fixed is the re-writing rules. They are domain dependent and thus they have to be designed to domain specifications.

Emotional expressions are added to the sounds performed through the increase or decrease of pitch, volume and duration associated with the communication message. Each robot will have a voice that is one of the sound class terminals. This is simpler than using timbre, which is the musical distinction between voices and instruments. The simulation of timbre would complicate the design of the language unnecessarily at this stage.

5.1 Calculating Emotional States

Physical state, physical knowledge (i.e. perception), social/physical interaction rules, and previous emotional state should be taken in consideration [7] in calculating an emotional state. Thus, calculating emotional state can be presented in an abstract form as follows:

$$eS_{n+1} \Leftarrow \lambda_e eS_n^{E_{\mathbb{R}}, B_{\mathbb{R}}} + \lambda_{ph} S_{n+1}^{B_{\mathbb{R}}, P_{\mathbb{R}}} \quad (11)$$

Where $\mathbb{R}$ is the universal rule set of interaction. B represents the belief model using the universal set of rules and P is the perception of the physical environment. λ is a discount function operates under the influence of the dominant emotion and physical perception respectively. The actual implementation of these function can vary of course from simplistic form to a complex adaptive functions. For example, λ can be a fixed number or non-dertministic function that calculates the discount value according to the current perceived emotional and physical states.

$$\forall e : Emotion \cdot \forall s : State \cdot Map(s)_{emotion \cdot rules} \rightarrow \langle SP, SD, SV \rangle \quad (12)$$

Formula 12 provides the mapping between the emotions triggered by state stimuli and the sound message elements of SP, SD and SV under emotional rules as they are defined in a given application design (see section 5).

5.2 Emotion External Modeling: Sound Expression

The external representation of emotions are usually done through demonstrative physical behavior, e.g. [6, 22]. In fact, what is often being expressed is the impact of emotions on the behavior selection without explicit expression of emotions. In uttering sounds for the purpose of communication both behavioral and emotional content being delivered implying the reasoning and personality behind the message. The emotional content in the language proposed will be delivered through the sound properties to be uttered. These properties are contained as part of the structural phonetic being communicated as defined by the application rules, emotions internal modeling, and MLEIR specifications.

5.3 Guidelines for Re-Writing Rules

Re-writing rules concerns the production and recognition of sound within the musical context defined by the language and the domain. A distinction needs to be made between the rules which will be dealing with the structures on one hand and sound recognition and production, as a form of signal processing, on the other. Sound production rules are the easiest to implement. These are particularly useful for human-computer interaction. Having an oscillator device or accessible sound production module in the machine, e.g. sound card, is all that is needed to produce the sound equivalence of inferred structures. Sound recognition is slightly more complex since it has to be done on two stages. First, a signal processing module is required to extract the sound features. Then the re-writing rules should map these sound blocks based on their features to structural containers based on the MLEIR. Finally, emotions and emotional state need to be reflected in re-writing rules. The complexity of emotional representation and calculation should suite the application and deployment device. Complex emotion estimation functions will not suite reactive limited resources robots whilst some complexity can be afforded on advance PDA that supports Java programs.

5.4 Case Study: a Dog's Sound-based Behavioral Patterns

By observing Alex, a Springer Spaniel dog, some repetitive sound-based behavioral patterns can be identified. These patterns are the basis of the emotional rules proposed here to link between sound produced (or to be produced) and basic emotions driven from feelings.

The first noticeable pattern is the short high pitch / high volume sound he produces in response of sudden pain or danger. Rule 5.4.1 represents this pattern. We use 'Sudden' and 'Pain' to avoid a hard concretized design at this stage. The meaning of descriptors is left to interpretation during implementation.

Rule 5.4.1 (Urgency)

$$\begin{aligned} \exists \alpha : \Gamma \cdot \forall Sr \in Sensors_{\alpha} \cdot \exists \kappa : \Sigma \\ \cdot Sr_{\kappa} \rightharpoonup' Sudden \cdot Pain' \Longrightarrow \\ (Suspends_{\alpha \cdot actions} \wedge Urgent_{SM})_{t_0 \succ t_n \leq cycle} \mapsto \\ SM \triangleleft SD =' Short' SP =' High' \wedge SV =' High' \end{aligned}$$

The second noticeable sound pattern is the sound of stress that is usually prolonged low or high pitched sound with usually high volume. In fact the volume increase in relation to the pain almost following a Gaussian function. Rule 5.4.2 attempts to capture that. The appearance of $\vee$ operator in this rule complicates the implementation of action selection (i.e. inference engine).

Rule 5.4.2 (Stress)

$$\begin{aligned} \exists \alpha : \Gamma \cdot \forall Sr \in Sensors_{\alpha} \cdot \exists \kappa : \Sigma \\ \cdot Sr_{\kappa} \rightharpoonup' Long \cdot Pain' \Longrightarrow \\ *(Suspends_{\alpha \cdot actions} \wedge Stress_{SM})_{t_0 \succ t_n \leq cycle} \mapsto \\ SM \triangleleft SD =' Long' \wedge SP =' Low \vee \\ High' \wedge SV =' High' \end{aligned}$$

The third noticeable sound patten is the lazy voice of calmness. It is often a stretched out voice of low volume and low pitch. Rule 5.4.3 represents this pattern in response to a 'Long Comfort'.

Rule 5.4.3 (Calm)

$$\begin{aligned} \exists \alpha : \Gamma \cdot \forall Sr \in Sensors_{\alpha} \cdot \exists \kappa : \Sigma \\ \cdot Sr_{\kappa} \rightharpoonup' Long \cdot Comfort' \Longrightarrow \\ *(Suspends_{\alpha \cdot actions} \wedge Calm_{SM})_{t_0 \succ t_n \leq cycle} \mapsto \\ SM \triangleleft SD =' Long' \wedge SP =' Low' \wedge SV =' Low' \end{aligned}$$

Excitement is often achieved by a treat or by coming home. This may be seen as a 'Sudden Comfort' of having a treat. After a while, it wears off and Alex goes back into a 'Calm' mood. There is usually an associated body gesture of tail wagging, which is outside the scope of this research; even though it is important to be considered in developing a full animal-like creature.

Rule 5.4.4 (Excited)

$$\begin{aligned}\exists \alpha : \Gamma \cdot \forall Sr \in Sensors_{\alpha} \cdot \exists \kappa : \Sigma \\ \cdot Sr_{\kappa} \rightharpoonup' Sudden \cdot Comfort' \Longrightarrow \\ *\left(Suspends_{\alpha \cdot actions} \wedge Excited_{SM}\right)_{t_0 \succ t_n \leq cycle} \mapsto \\ SM \lhd SD =' Short' \wedge SP =' Normal_{often} \vee \\ High' \vee SV =' High_{often} \vee Normal'\end{aligned}$$

Fear is a slightly more complex emotion. Based on observation, there are at least two reasons that generate fear feelings. The first one $(5.4.5)$ is the fear that is based on the unknown. Something that is new and seems threatening leads the dog into a cautious mood. The second $(5.4.6)$ is fear from remembering a bad experience that had happened in the recent past that led to discomfort.

Rule 5.4.5 (Fear of Unknown)

$$\begin{aligned}\exists \alpha : \Gamma \cdot \forall Sr \in Sensors_{\alpha} \cdot \exists \kappa : \Sigma \\ \cdot Sr_{\kappa} \rightharpoonup' Unknown'_{\rightharpoonup(stress \wedge urgency)} \Longrightarrow \\ *\left(Suspends_{\alpha \cdot actions} \wedge Fear_{SM}\right)_{t_0 \succ t_n \leq cycle} \mapsto \\ SM \lhd SP =' Low' \wedge SV =' High'\end{aligned}$$

Rule 5.4.6 (Fear from Memory)

$$\begin{aligned}\exists \alpha : \Gamma \cdot \forall Sr \in Sensors_{\alpha} \cdot \exists \kappa : \Sigma \\ \cdot Sr_{\kappa} \rightharpoonup \\ 'mem\left(stress \wedge urgency\right)' \Longrightarrow \\ *\left(Suspends_{\alpha \cdot actions} \wedge Fear_{SM}\right)_{t_0 \succ t_n \leq cycle} \mapsto \\ SM \lhd SP =' Low' \wedge SV =' High'\end{aligned}$$

Rule 5.4.6 gives a good example of emotions interrelations where a combination of stress and urgency crystallize into fear.

6 Analytical Discussion

By means of a case study, some of the properties of the generic language MLEIR were demonstrated. The first property is extendibility. The language provides the building blocks for varying structures of expressions. The generic form of re-writing rules enables a flexible generation of equivalent phonetics of the expression structures uttered. The second property is the need of substitution for practical implementation. The language is very generic that in its current form it only provides a framework for implementation. Guidelines for re-writing rules to enable the development of meaningful rules of given application was presented. However, there are no verification guard process. In addition, there are no specification guidelines for specific functions beyond the generic emotions and belief modeling. The use of Z notation for specification, however, enables us to import from the rich developments in specification and verification field [42]. This is still to be formalized in the framework of MLEIR.

There are many extensions that can be applied to MLEIR. For example, the language syntax could be extended by introducing chords, sharp and flat nodes. Similarly, the grammar could be extended by introducing melodic sequences and compositional rules. The fact that we refer to sound classes is meant to make the link to serial composition techniques for which computational models can be developed to enable evolvable phonetics [8]. Finally, the range of emotions could be extended that can be presented within this language. This will require farther research in psychology of emotions theories [47], animal intelligence and the semantics of the musical language developed. In addition, the implementation of the language can go beyond the simplified version of rules to include: learning capabilities, a neural model of MLEIR and its extended models include using layered cognitive maps. This will enable the signal processing of sound and the association between signals and their alphabetical structure equivalent such the case between sound of words and their symbolic representation. In fact, we have proven that neural networks can be used in natural language processing [1] thus opening the door to further extensions.

7 Implementation of MLEIR based Lego Society

The language is implemented on Lego robots which have several limitations. The Lego robots do not have sound input thus it is difficult to have sound recognition. Lego robot has limited range of musical sounds it can produce. Limited sensory and processing power which goes well with the assumption of reactive agents and primitive animal behavior. Thus testing is adapted to accommodate these limitations by making number of assumptions in implementing a sub-set of MLEIR. It is assumed that there is no memory based emotions. In other words, the robot does not learn, instead it reacts based on individual states without historical data. Each robot broadcasts its sound messages using the Infrared port on the Lego RCX instead of sound card. Figure 1 shows the Lego model used. There are two touch sensors, one light sensor, and the Infrared port is at the front of the micro-controller box on the top.

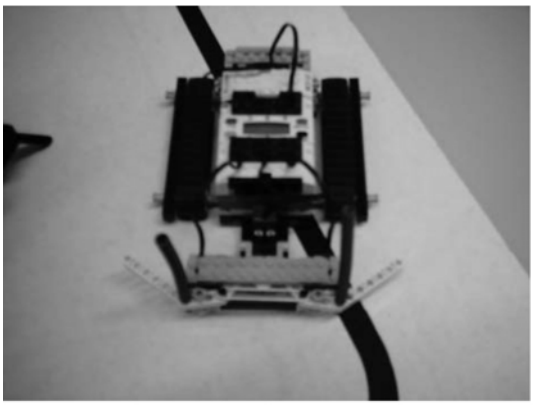


Figure 1: Lego Mind Storm Model

Each robot will response to the first message it receives. In implementing behaviors, there are three primitive actions the robot can take: forward, left turn and

Algorithm 1 Own behavior

```
while (true) do
  case Calm:
  move randomly;
  produce calm messages;
  check sensors;
  if sensors are clear then
    continue;
  else
    break;
  end if
  endcase Calm
  case Urgency:
  if touch sensors are active then
    move away fast;
    produce Urgency message;
  end if
  if touch sensors are clear then
    break;
  end if
  endcase Urgency
  case Excited:
  if light sensor is active then
    repeat
      go around in circle ;
      produce Excited messages;
    until eating_time = 0;
  end if
  clear light sensor;
  break;
  endcase Excited
end while
```

Algorithm 2 Communication response

```
while (true) do
  case Fear:
  check for messages;
  if other[i].message = 'urgent' then
    repeat
      turn and move away;
      produce afraid messages
    until no other[].message = 'urgent';
  else
    break;
  end if
  endcase Fear
  case Stress:
  check for messages;
  if other[i].message = 'excited' then
    repeat
      move forward;
      produce stress messages and self-naming messages
    until no other[].message = 'excited' or other[i].message = 'acknowledge';
  end if
  endcase Stress
  defult:
  if other[i].message ='acknowledge' then
    become excited
  else
    break;
  end if
  enddefult
end while
```

right turn. Backward is unnatural and unnecessary action, which effect can be emulated using a combination of turns. The behaviors are implemented into parallel tasks that starts and stops based on the emotional rules. Thus mental-emotional states defined are: calm state, urgency state, fear state, stress state and excited state. The sound messages that are associated with these rules are generated as the robot moves from one state, i.e. task, to another. Algorithm 1 shows action selection of own responses. In this case, there is non-verbal expression of emotions and thoughts.Algorithm 2 shows the communication response that the robot will express. In this case, the behavior comes in the form of communicative utterance that will have impact on other robots.

8 Conclusion

A Musical Language for Emotional Interaction ($MLEIR$), which supports robot-robot and robot-human interaction, has been proposed in this paper. The syntactical definition of the language was provided in generic form with operational functions to facilitate communication and emotional expression. The aim is to develop a language that has simple structure which is expandable, computationally inexpensive, and more natural to machines than human natural languages with structural format enable translation and phonetic evolution. The working of this language was demonstrated through to contrasting examples. First example, shows how this language may be implemented in swarm collective of limited resources robots to enable more realistic animal like interaction. Second example showed how it can be used in user's interaction with relatively limited resources devices such as mobile phones. Analytical discussion provided a preliminary assessment of the musical language (MLEIR) and its uses. Further extensions and studies will be required to explore the full potential of the proposed language.

References

[1] Hasan M. Alserhan and Aladdin S. Ayesh. An application of neural network for extracting arabic word roots. *WSEAS TRANSACTIONS on COMPUTERS*, Issue 11, Volume 5(ISSN 1109-2750):pp. 2623–2627, November 2006.

[2] James L. Alty and Dimitrios Rigas. Exploring the use of structured musical stimuli to communicate simple diagrams: the role of context. *International Journal of Human-Computer Studies*, 62(1):21–40, January 2005.

[3] J. L. Arcos, R. L. de Mantaras, and X. Serra. Saxex: a case-based reasoning system for generating expressive musical performances. In P. R. Cook, editor, *Proceedings of the International Computer Music Conference, Computer Music Association*, pages 329–336, 1997.

[4] Phuoc Nguyen Hao Tang Asad Yousuf, William Lehman. Animatronics and emotional face displays of robots. Online, 2003.

[5] A. Ayesh. Self reference in ai. Master's thesis, Computer Science Dept, Essex University, 1995.

[6] A. Ayesh. Emotionally motivated reinforcement learning based controller. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, volume 1, pages 874–878 vol.1, 2004.

[7] A. Ayesh and J. Cowell. Emotional analysis of facial expressions. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, volume 1, pages 672–676 vol.1, 2004.

[8] Aladdin Ayesh and Andrew Hugill. Genetic approaches for evolving form in musical composition. In *Proceddings of the 23rd IASTED International Multi-Conference Artificial Intelligence and Applications (AIA2005)*, pages 318–321, 2005.

[9] Tucker Balch and Ronald C. Arkin. Communication in reactive multiagent robotic systems. *Autonomous Robots*, 1(1):27–52, 1994.

[10] Barbara Becker. Social robots Ð emotional agents: Some remarks on naturalizing man-machine interaction. *International Review of Information Ethics*, 6:37–45, 2006.

[11] P. Bellini, J. Barthelemy, P. Nesi, and G. Zoia. A proposal for the integration of symbolic music notation into multimedia frameworks. In *Web Delivering of Music, 2004. WEDELMUSIC 2004. Proceedings of the Fourth International Conference on*, pages 36–43, 2004.

[12] Wallace Berry. *Structural Functions in Music*. Dover Publications, Inc., New York, 1987.

[13] William Blewitt, Aladdin Ayesh, Robert I John, and Simon Coupland. A millenson-based approach to emotion modelling. In *Conference on Human System Interaction (HSI'08)*, 2008.

[14] Rodney Brooks. Intelligence without reason. In *Proceedings of the 12th International Joint Conference on Artificial Intelligence (IJCAI-91)*, 1991.

[15] Dolores Canamero. Issues in the design of emotional agents. In *AAAI Fall Symposium on Emotional and Intelligent: The tangled knot of cognition, AAAI Press, Menlo Park, CA.*, pages 49–54, 1998.

[16] Bee Yong Chua and Guojun Lu. Perceptual rhythm determination of music signal for emotion-based classification. In *Multi-Media Modelling Conference Proceedings, 2006 12th International*, page 8pp., 4-6 Jan. 2006.

[17] Nicholas Cook. *A Guide to Musical Analysis*. J. M. Dent & Sons Ltd, London, 1987.

[18] E. Coutinho, M. Gimenes, J. M. Martins, and E. R. Miranda. Computational musicology: An artificial life approach. In *Proc. portuguese conference on Artificial intelligence epia 2005*, pages 85–93, 2005.

[19] M. Dragone, B.R. Duffy, and G.M.P. O'Hare. Social interaction between robots, avatars & humans. In *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*, pages 24–29, 13-15 Aug. 2005.

[20] U. Esnaola and T. Smithers. Mirela: a musical robot. In *Computational Intelligence in Robotics and Automation, 2005. CIRA 2005. Proceedings. 2005 IEEE International Symposium on*, pages 67–72, 27-30 June 2005.

[21] S. Gadanho. *Reinforcement Learning in Autonomous Robots: An Empirical Investigation of the Role of Emotions*. PhD thesis, University of Edinburgh, 1999.

[22] Sandra Clara Gadanho. Learning behavior-selection by emotions and cognition in a multi-goal robot task. *J. Mach. Learn. Res.*, 4:385–412, 2003.

[23] L. Gupta and Suwei Ma. Gesture-based interaction and communication: automated classification of hand gesture contours. 31(1):114–120, 2001.

[24] Bon-Woo Hwang, Sungmin Kim, and Seong-Whan Lee. A full-body gesture database for automatic gesture recognition. In *Proc. 7th International Conference on Automatic Face and Gesture Recognition FGR 2006*, pages 243–248, 2006.

[25] M. Ihara, N. Watanabe, and K. Nishimura. A gesture description model based on synthesizing fundamental gestures. In *Proc. IEEE Southeastcon '99*, pages 47–52, 1999.

[26] ISO/IEC. Extended bnf. *http://www.cl.cam.ac.uk/ mgk25/iso-14977.pdf*, 1996.

[27] Patrik N. Juslin. Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance*, 26(6):1797–1812, December 2000.

[28] K. Kahol, P. Tripathi, and S. Panchanathan. Computational analysis of mannerism gestures. In *Proc. 17th International Conference on Pattern Recognition ICPR 2004*, volume 3, pages 946–949 Vol.3, 2004.

[29] T. Kirishima, K. Sato, H. Narita, and K. Chihara. Realtime gesture recognition under the multi-layered parallel recognition framework of qvips. In *Proc. Third IEEE International Conference on Automatic Face and Gesture Recognition*, pages 579–584, 1998.

[30] S. K. Korde and K. C. Jondhale. Hand gesture recognition system using standard fuzzy c-means algorithm for recognizing hand gesture with angle variations for unsupervised users. In *Proc. First International Conference on Emerging Trends in Engineering and Technology ICETET '08*, pages 681–685, 2008.

[31] D.E. Kroodsma and E.H. Miller. *Acoustic Communication in Birds*. New York: Academic Press, 1982.

[32] Dong-Soo Kwon, Yoon Keun Kwak, J. C. Park, Myung Jin Chung, Eun-Sook Jee, Kyung-Sook Park, Hyoung-Rock Kim, Young-Min Kim, Jong-Chan Park, Eun Ho Kim, Kyung Hak Hyun, Hye-Jin Min, Hui Sung Lee, Jeong Woo Park, Su Hun Jo, Soon-Young Park, and Kyung-Won Lee. Emotion interaction system for a service robot. In *Proc. 16th IEEE International Symposium on Robot and Human interactive Communication RO-MAN 2007*, pages 351–356, 2007.

[33] A.S. Lampropoulos, D.N. Sotiropoulos, and G.A. Tsihrintzis. Individualization of music similarity perception via feature subset selection. In *Systems, Man and Cybernetics, 2004 IEEE International Conference on*, volume 1, pages 552–556vol.1, 10-13 Oct. 2004.

[34] J. C. Lementec and P. Bajcsy. Recognition of arm gestures using multiple orientation sensors: gesture classification. In *Proc. 7th International IEEE Conference on Intelligent Transportation Systems*, pages 965–970, 2004.

[35] Tao Li and M. Ogihara. Toward intelligent music information retrieval. *Multimedia, IEEE Transactions on*, 8(3):564–574, June 2006.

[36] Lie Lu, D. Liu, and Hong-Jiang Zhang. Automatic mood detection and tracking of music audio signals. *Audio, Speech and Language Processing, IEEE Transactions on [see also Speech and Audio Processing, IEEE Transactions on]*, 14(1):5–18, Jan. 2006.

[37] J. McCormack. Grammar based music composition. 2004.

[38] E. R. Miranda. The art of rendering sounds from emergent behaviour: cellular automata granular synthesis. In *Proc. 26th Euromicro Conference*, volume 2, pages 350–355 vol.2, 2000.

[39] Jean-Sebastien Monzani and Daniel Thalmann. A sound propagation model for interagents communication. In *Virtual Worlds Second International Conference, VW 2000, Paris, France*, pages 135–146, 2000.

[40] Wang Muyuan, Zhang Naiyao, and Zhu Hancheng. User-adaptive music emotion recognition. In *Signal Processing, 2004. Proceedings. ICSP '04. 2004 7th International Conference on*, volume 2, pages 1352–1355vol.2, 31 Aug.-4 Sept. 2004.

[41] A. Nishikawa, A. Ohnisi, and F. Miyazaki. Description and recognition of human gestures based on the transition of curvature from motion images. In *Proc. Third IEEE International Conference on Automatic Face and Gesture Recognition*, pages 552–557, 1998.

[42] Ben Potter, Jane Sinclair, and David Till. *An Introduction to Formal Specification and Z*. Prentice Hall, 1991.

[43] Judy Robertson. Real-time music generation for a virtual environment. 2004.

[44] B. Robins and K. Dautenhahn. The role of the experimenter in hri research: A case study evaluation of children with autism interacting with a robotic toy. In *Proc. 15th IEEE International Symposium on Robot and Human Interactive Communication RO-MAN 2006*, pages 646–651, 2006.

[45] B. Robins, N. Otero, E. Ferrari, and K. Dautenhahn. Eliciting requirements for a robotic toy for children with autism - results from user panels. In *Proc. 16th IEEE International Symposium on Robot and Human interactive Communication RO-MAN 2007*, pages 101–106, 2007.

[46] Klaus R. Scherer. Expression of emotion in voice and music. *Journal of Voice*, 9(3):235–248, September 1995.

[47] K. T. Strongman. *The Psychology of Emotion: Theories of Emotion in Perspective*. Chichester: John Wiley & Sons, 2000.

[48] Hao Tang, Yun Fu, Jilin Tu, M. Hasegawa-Johnson, and T. S. Huang. Humanoid audio and visual avatar with emotive text-to-speech synthesis. 10(6):969–981, 2008.

[49] Hao Tang, Yun Fu, Jilin Tu, T. S. Huang, and M. Hasegawa-Johnson. Eava: A 3d emotive audio-visual avatar. In *Proc. IEEE Workshop on Applications of Computer Vision WACV 2008*, pages 1–6, 2008.

[50] Paul Vickers and James L. Alty. Musical program auralisation: a structured approach to motif design. *Interacting with Computers*, 14(5):457–485, October 2002.

[51] Paul Vickers and James L. Alty. Using music to communicate computing information. *Interacting with Computers*, 14(5):435–456, October 2002.

[52] Stewart W. Wilson. The animat path to AI. In Jean-Arcady Meyer and Stewart W. Wilson, editors, *From animals to animats*, pages 15–21, 1991.

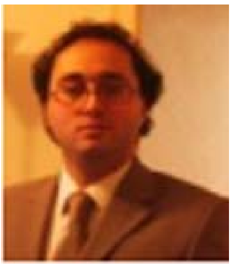
[53] Qijun Zhang and E. R. Miranda. Towards an evolution model of expressive music performance. In *Proc. Sixth International Conference on Intelligent Systems Design and Applications ISDA '06*, volume 2, pages 1189–1194, 2006.

[54] Yongmian Zhang and Qiang Ji. Active and dynamic information fusion for facial expression understanding from image sequences. 27(5):699–714, 2005.

[55] Wenming Zheng, Xiaoyan Zhou, Cairong Zou, and Li Zhao. Facial expression recognition using kernel canonical correlation analysis (kcca). 17(1):233–238, 2006.

[56] Yongwei Zhu and M.S. Kankanhalli. Precise pitch profile feature extraction from musical audio for key detection. *Multimedia, IEEE Transactions on*, 8(3):575–584, June 2006.

Biography



I am currently lecturing mainly on intelligent robotics, multi-agent systems, and multimedia related topics (games and entertainment computing). I am also a software architect and consultant. I am very interested in emerging technologies such as object-oriented systems, agents, neuro and adaptive systems, quantum devices and nanotechnology. My research is in the area of Emotions and Social Swarms and focuses on cognition, natural languages, logic and reasoning about knowledge and behavior with applications in cognitive agents/robots, sensor networks, data fusion and mining, and creative computing (Recent Papers or Publications). I am currently the group coordinator of Intelligent Mobile Robots and Creative Computing. In addition, I maintain a resource pages related to the Grand Challenge on Architectures for Brain and Mind.