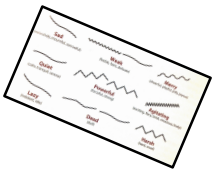
Could a robot be moved by music?

Using Motherese to Develop Multimodal Emotional Intelligence for Robots



Angelica Lim, Ph.D. Candidate, Okuno Lab, Department of Intelligence Science and Technology, Kyoto University

Goal A robot that develops multimodal emotional intelligence as < 1 yr olds

1. Background

In 2010, we developed robots that could play music with humans.

[Lim et al. 2010]

Expression
GERMS: Generative rules, Emotion, Randomness, Motion constraints, Stylistic unexpectedness
[Juslin 2003]

Emotion
Display
Recognition
Authenticity?

While useful for practice, these and other empathetic companion robots are not accepted because they do not express true feelings: "nobody likes fake emotion."

2. Research questions

How can we create an emotionally authentic robot?

Authentic: "Made or done the same way as an original" *(Merriam Webster)*

Can we make a robot develop and learn emotions like a human?

3. Development of emotional intelligence

Expression

First appearance	Emotion
Primary Emotions	
3 months	Joy Sadness Disgust
2 to 6 months	Anger
First 6 months	Surprise
6 to 8 months	Fear (peaks at 18 months)
Self-Conscious Emotions	
1½ to 2 years	Empathy Jealousy Embarrassment

[Sigelman & Rider, 2012]

Recognition

7-month old infants:
- can recognize happy/angry **voices** [Grossman et al, 2006]
- can recognize happy/angry **motions** (point-light display) [Soken and Pick, 2002]

9-month old infants:
- can discriminate sad/happy **music** [Flom et al., 2008]

4. Scope of my research

- **an emotion model developed through interaction** with a human, similar to <1 year olds
- **expression** and **recognition** of emotion in **multiple modalities**
- **primary emotions** happiness, sadness, anger, fear

5. Proposed Approach

Motherese: "Concurrent with the exaggerated speech of motherese, there are probably exaggerated facial displays. Child-centered displays may serve as opportunities for **learning about affective events.**" [Soken and Pick, 1992]

6. SIRE Emotion Model

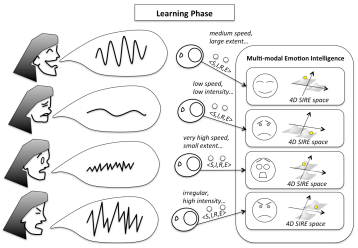
It's long been suggested that music emotions have its origins in the voice [Juslin & Laukka, 2001]. In 2011, we developed a cross-modality model to describe happiness, sadness, anger and fear through 4 common perceptual parameters: **speed, intensity, regularity and extent (SIRE)** [Lim et al, 2012]

		Modality mappings to relevant emotional features		
Parameter	Description	Voice	Gesture	Music
Speed	slow vs. fast	*speech rate* [20], pauses [18]	*velocity* [35], animation [36], quantity of motion [28]	*tempo* [24,26]
Intensity	gradual vs. abrupt	*voice onset rapidity* [18], articulation [20]	*acceleration* [35], power [37]	*note attack* [24], articulation [26]
Regularity	smooth vs. rough	*jitter* [18], voice quality [18,20]	directness [35], *phase shift* [38,39], fluidity [27]	*microstructural irregularity* [26], timbral roughness [24]
Extent	small vs. large	*pitch range* [20], *loudness* [18]	*spatial expansiveness* [36,37], contraction index [35]	*volume* [24,26]

Features in *italics* are used in our experiments.

7. Training by voice, Testing of motion

We train 2-mix Gaussian Mixture models with normalized SIRE samples from voice, one for each of our 4 emotions.



Recognition of emotion by taking the model with max probability.
Expression of emotion by sampling the desired GMM and perceptually mapping the SIRE to **music** [Lim et al, 2012] or **motion** [Lim et al, 2011].

8. Experiments

Recognizing emotion in motion (gait)

Training: Voice

Berlin Emotional Speech Database

- Wave files
- 10 subjects
- **Training**: 408 voice samples

Testing: Gait

Body Movement Library

- Feet positions
- 28 subjects
- **Testing**: 236 gait samples

- Happy and sad walks recognized well over chance
- Hot anger mis-recognized as elation
- Anxious fear confused as sadness

[Lim and Okuno, 2012]

Expressing emotion in voice & motion (see video)

- Happiness and sadness again well recognized.
- Terror fear expressed approaching the object was mistaken as happiness
- Cold anger expressed approaching the object lacked dominance component [Lim and Okuno, in prep.]

9. Future Work

- Offline → Online training
- Happy/sad as basis, scaffolding for anger and fear
- Grounding learned SIRE samples in positive and negative somatosensory states, e.g. battery levels [Ask to learn more!]