# Expressive Speech Bubbles

MAS 630 Affective Computing
Final Project

10th December 2013

Ben Bloomberg & Pip Mothersill

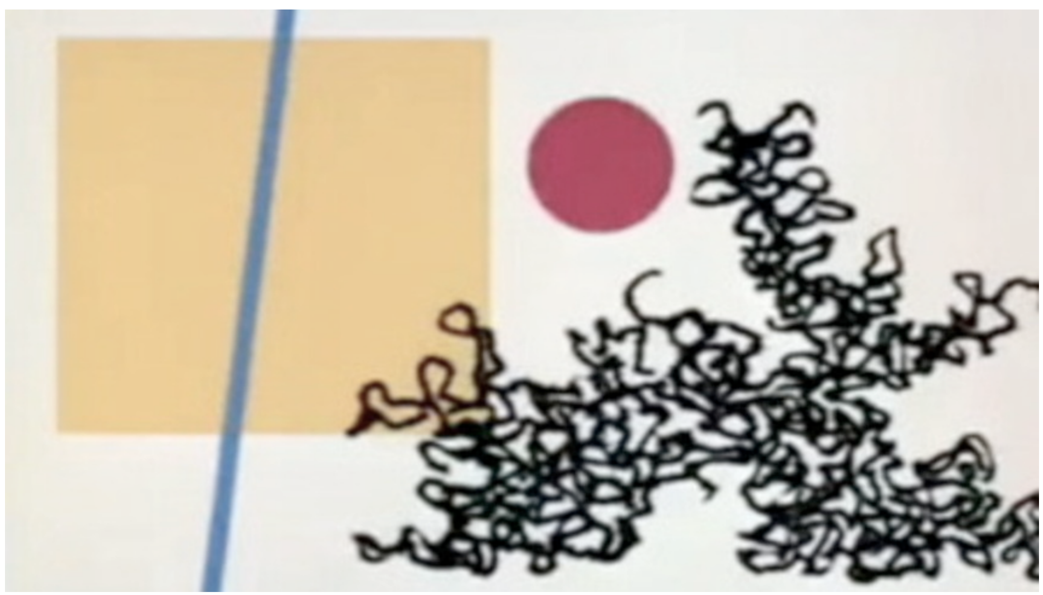
Please view in Slide Show – otherwise you'll miss the pretty animations!

# Introduction

# Project Summary

We can often infer emotional expression from very simple shapes and sounds; from a simple line leaning 'forward' expressing dynamism and engagement towards another shape (Juster), to elementary notes moving up and down a scale to evoke positive or negative feelings. The combination of these emotive shapes and sounds can then become an enormously communicative medium for emotional expression, through even the most simple and abstract stimuli. This union of simple shapes and sounds can evoke very powerfully emotive stories, and is beautifully demonstrated in the short movie "The Dot and the Line: A Romance in Lower Mathematics" by Chuck Jones.

This project is an exploration into how various shapes and sounds can be used to communicate emotive expressions, and how these shapes and sounds can be used to help people create more emotive stories.



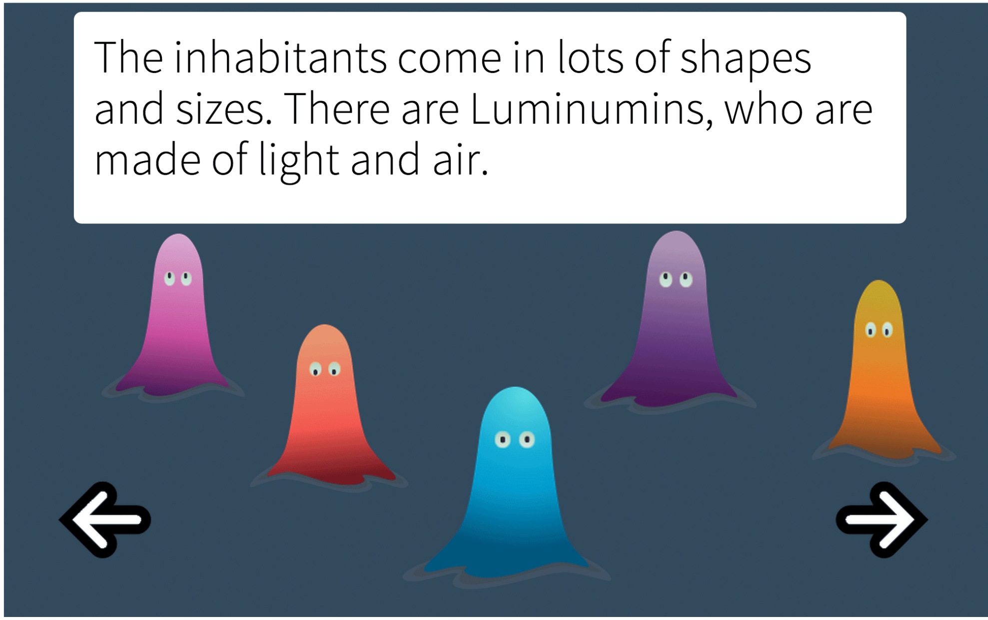
Juster, N. (2000). The dot and the line: A romance in lower mathematics. Chronicle Books.

# Project Application

Considering the huge potential of emotional expression that simple shapes and sounds can have, a relevant application of this research was in some sort of interactive story telling application. In this project, we were inspired by the StoryScape application developed by Micah Eckhardt in the MIT Media Lab Affective Computing group. This is an online illustrated storytelling app which allows the writers to create “highly interactive and customizable stories” to share with their readers (Eckhardt).

The platform allows the integration of simple animations created from many static images played in succession for differing lengths of time (gifs), as well as the recording and playing of different sounds. An example of some of the animations possible can be seen in the animation below. Using this application as inspiration, we will aim to create animated emotive shapes and sounds for objects which users of the platform can use in their own stories to add emotional meaning to them.


Eckhardt M., Goodwin M., Picard R. W., "StoryScape: A Social Illustrated Primer", in the Extended Abstract of IMFAR 2013, San Sebastian, Spain, May 2-4, 2013
http://affect.media.mit.edu/projects.php?id=3480

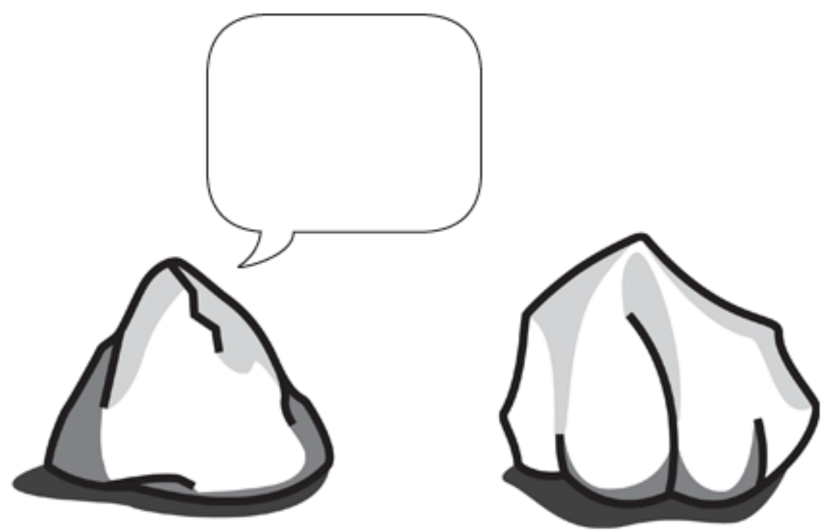



An example animation from one of the stories in StoryScape (courtesy Micah Eckhardt)
http://afc.media.mit.edu/stories/browse/

# Project Application

But what shape should we create? We wanted to create some shapes and sounds that would be generally applicable to any of the stories users of StoryScape wanted to create. As users in StoryScape can upload any images they wanted, this made the task of selecting a shape or object which would be broadly useful to all the types of stories that the users wanted to create. Researching existing stories created on the platform highlighted the one commonality across all the stories however: speech bubbles to contain the narration or dialogue text. The existing speech bubbles were very simple boxes, and if the user chose not to upload any sounds or voices relating to the speech bubble, it is only the text contained within the bubble that expresses the emotional sentiment of the story. What if the speech bubbles themselves could express emotion? How could their shape change to indicate the emotion of the text contained within? How could emotional sounds attached to the speech bubble enhance the emotional communication of the message?

So to explore this idea of adding emotional communication to the text areas in the StoryScape app, we decided to create a range of **expressive speech bubbles** – just like the one below!



# Project Aims

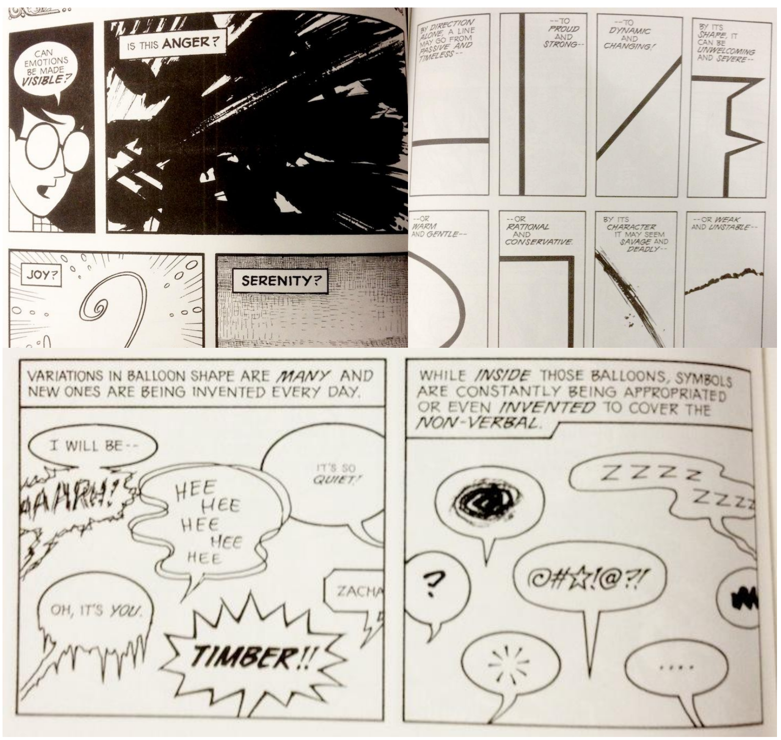
Given our end goal of developing a range of emotive shapes and sounds for speech bubbles to enhance the communication of emotion in the text areas of the StoryScape app, the aims for this project included:

- Investigate the representation of emotion through sound and shape
- Develop hypothetical frameworks for emotive sounds and shapes applied to expressive speech bubbles
- Evaluate the effectiveness of emotive sounds and shapes in qualitative user studies

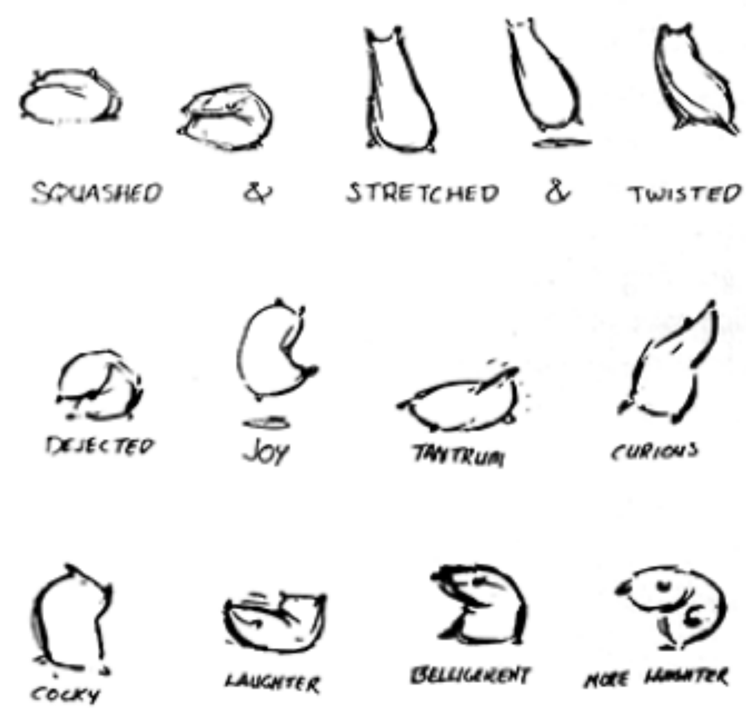
# Background

# Background: Emotive Shapes

Emotions can be communicated through simple static or dynamic shapes. This has been shown extensively in the arts, from animation to comic books. Disney shows how even static images of a simple flour sack can show a range of emotion by adding anthropomorphic shapes and postures (Thomas). The different qualities of movement afforded by the dynamic medium of animation enhance this emotional communication even further. Comic book design also provides interesting examples of how even the most simple lines can communicate different emotional qualities, just through its smoothness or the angle of it on the page (Manning). Speech bubbles are also extensively used in comic book design, and can also provide some inspiration for how we can approach this area of expressive speech bubble shape in this project.



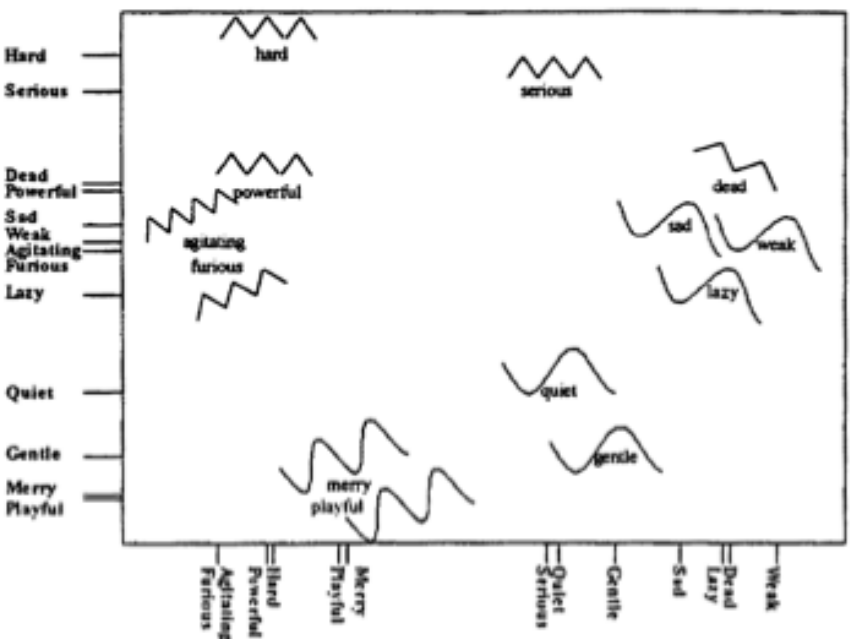

Manning, A. D. (1998). Understanding comics: The invisible art. *Professional Communication, IEEE Transactions on*, *41*(1), 66-69.



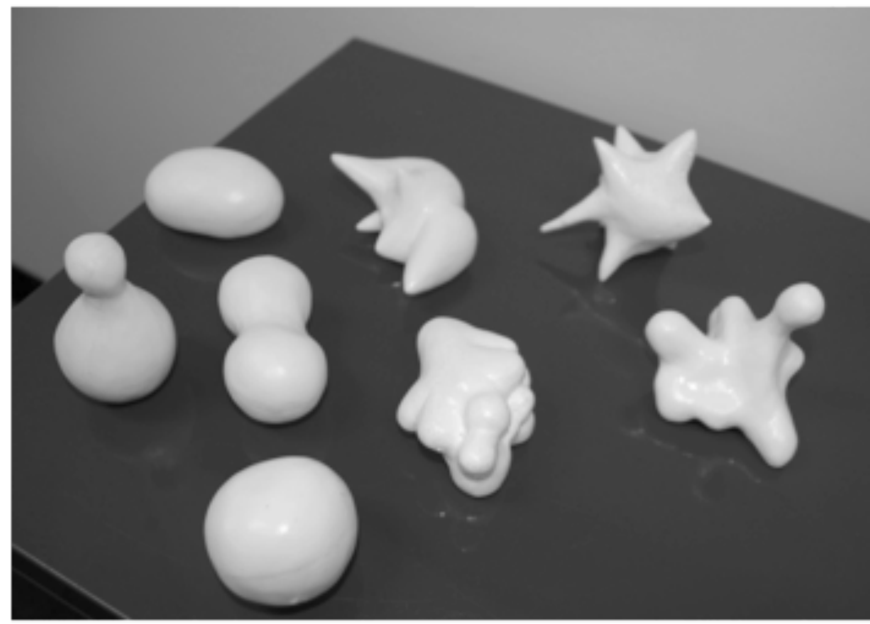

The animated flour sack from *"The Illusion of Life: Disney Animation"* by Johnston & Thomas, 1995.

# Background: Emotive Shapes

Moving from the emotive communication intuition of artists to more quantitative theories about our emotional perception of different shapes, there are studies that show there is actually some consistency in how we perceive certain shapes to emotions.  These experiments often use the emotional dimensions of Pleasure, Arousal and Dominance defined by Russell (Russell) and Mehrabian (Mehrabian) to map features of the lines to different emotions.  Simple lines with **angles leaning up were showed to convey more active emotions – i.e. higher Arousal and Dominance - and vice versa** (Poffenberger, Collier).  **Rounded and angular lines was also shown to relate to positive and negative Pleasure** (Isbister).




Poffenberger, A. T., & Barrows, B. E.  *"The Feeling Value of Lines"*  (1924)
Collier, G.L.  *"Affective synesthesia: Extracting emotion space from simple perceptual stimuli"*  (1996)

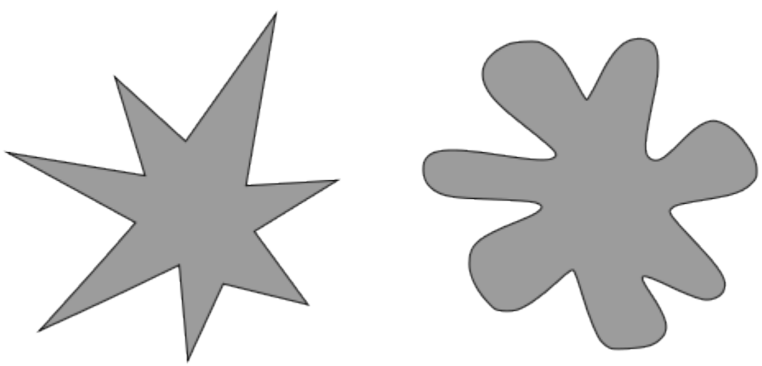


Isbister, K. et al.  *"The Sensual Evaluation Instrument: Developing an Affective Evaluation Tool"* (2006)

Russell, James A. A circumplex model of affect. *Journal of personality and social psychology* 39.6 (1980): 1161-1178
Mehrabian, A. (1996). Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in temperament. Current Psychology, 14(4), 261-292.

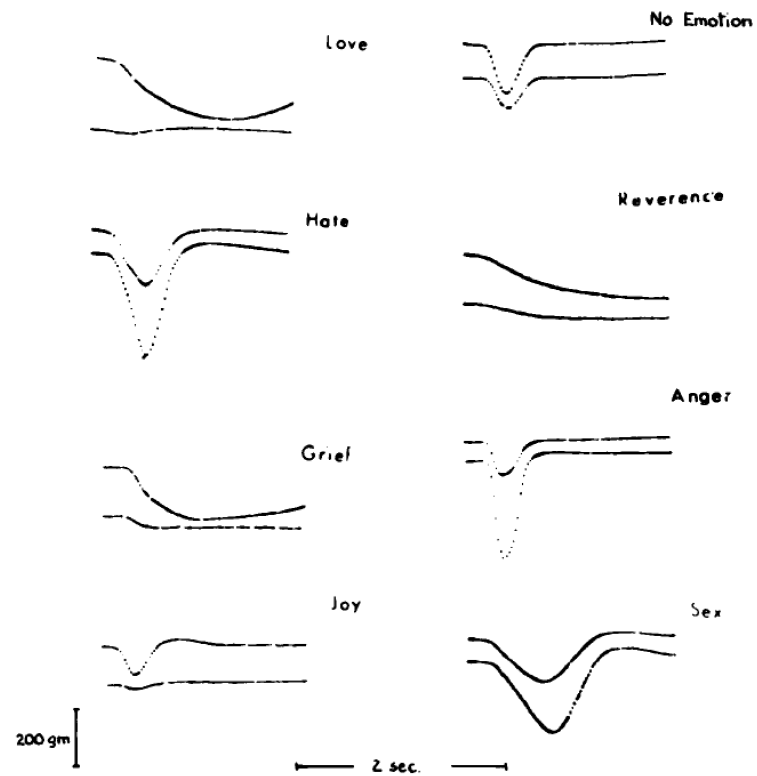
# Background: Emotive Shapes

There are also studies also that show there is some consistency in how we perceive certain shapes to sounds. A famous example of this confirmation of the existence of an underlying universal response to these emotionally expressive sensory stimuli is Clyne's work on describing our emotional response to different sounds – modes of responses he defined as 'sentic forms' (Clynes). Another well-known study is that of Kohler's 'maluma' and 'takete' experiment, where he showed that the majority of people matched the nonsense word 'malumba' to a line drawing of a smooth rounded shape and the word 'takete' to a spiky angular shape (Kohler).



Kiki Bouba

Kohler, W. (1947). *Gestalt psychology*. New York: Liveright.

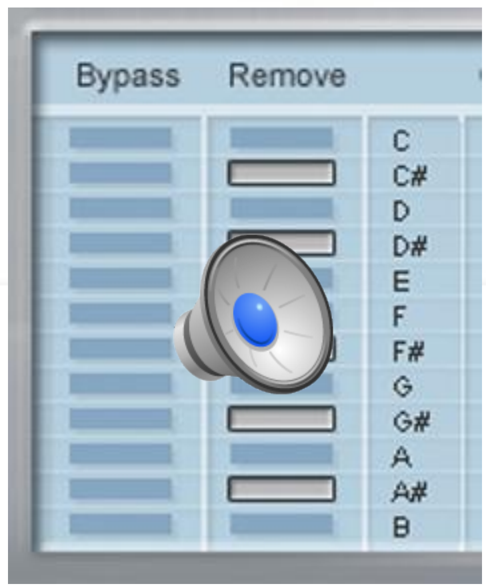



Clynes, M. (1973). Sentics: Biocybernetics of Emotion Communication *Annals of the New York Academy of Sciences*, *220*(3), 57-88.

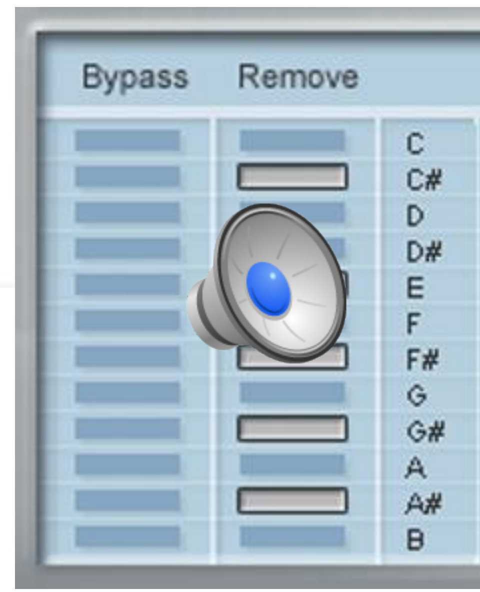
# Background: Emotive Sounds

Emotions can also be very widely communicated through different sounds.
Click on the images below to listen to some examples – what emotions do you think are being communicated?


Diatonic scale, ascending




Whole tone scale, ascending




Minor scale, descending

*(2010) Handbook of Music and Emotion*. ed. Juslin, P. N; Slobada, J. A. Oxford.
Meyer, L. (1961) *Emotion and Meaning in Music*. Chigaco, IL: University of Chicago Press.
*Hein (2013) Scales and Emotion with AutoTune http://www.ethanhein.com/wp/2010/scales-and-emotions/*

# Background: Emotive Sounds

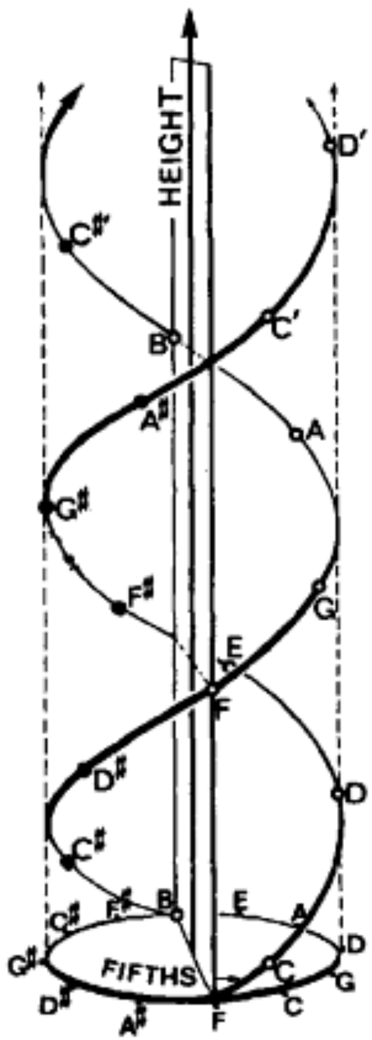
There is a vast amount of material connecting harmonic structure to emotion. The most basic and one of the earliest connections was that of Greek modes described by Plato and Aristotle. Certain relationships between scale degrees were said to induce specific emotions. In Aristotle's Politics, he states:

> *"...even in mere melodies there is an imitation of character, for the musical modes differ essentially from one another, and those who hear them are differently affected by each. Some of them make men sad and grave, like the so-called Mixolydian, others enfeeble the mind, like the relaxed modes, another, again, produces a moderate and settled temper, which appears to be the peculiar effect of the Dorian; the Phrygian inspires enthusiasm."*

Modern models, such as A Generative Theory of Tonal Music (1983) by Fred Lerdahl, consider tension and harmonic structure in terms of stability and instability. The stability is analyzed based on distance between scale degrees and the absolute distance from the scale's tonic. To similarly attempt to map these scales to the PAD emotion scales, we hypothesize in general that greater instability leads to higher arousal, lower dominance and lower valence. This is an application of Lerdahl's Tonal Hierarchy, which provides an atemporal analysis of harmonic space.  Once keyframes and scale directions are added to the mix, it is also possible to apply Lerdahl's Event Hierarchy. Since our combinations are computationally generated, we chose to create our initial hypothesis without considering time.


(2010) Handbook of Music and Emotion. ed. Juslin, P. N; Slobada, J. A. Oxford.
Meyer, L. (1961) Emotion and Meaning in Music. Chigaco, IL: University of Chicago Press.
Hein (2013) Scales and Emotion with AutoTune http://www.ethanhein.com/wp/2010/scales-and-emotions/
Aristotle (350 BCE) Politics, http://classics.mit.edu/Aristotle/politics.8.eight.html
Lerdahl (2001) Tonal Pitch Space. New York, NY; Oxford

# Background: Emotive Sounds




**Examples of note "distance" mapping**

Shepard's melodic map: A visualization of "distance" between notes using a double helix to shorten distance between octaves and fifths.

| | | | | | |
|---|---|---|---|---|---|
| A♯ | E♯ | B♯ | F𝄪 | C𝄪 | G𝄪 |
| F♯ | C♯ | G♯ | D♯ | A♯ | E♯ |
| D | A | E | B | F♯ | C♯ |
| B♭ | F | C | G | D | A |
| G♭ | D♭ | A♭ | E♭ | B♭ | F |
| E𝄫 | B𝄫 | F♭ | C♭ | G♭ | D𝄫 |

A lattice distance map: horizontal axis is perfect fifths and vertical axis is perfect thirds.

FIGURE 2.2 (a) Shepard's melodic map (Reproduced by permission of Academic Press); (b) Riemann's *Tonnetz*.

Lerdahl (2001) Tonal Pitch Space. New York, NY; Oxford

# Background: Emotive Sounds

**What does this mean for us?**

- In general, we'll hypothesize that **scales going up are less stable** and **scales returning to the tonic are more stable**.
- The **diatonic major scale**, comprised mainly of half and whole steps is the **most stable**. This corresponds to **positive valence.**
- **Heptatonic** and **whole-tone scales** contain intervals with greater harmonic tension and this makes them the **least stable**. This corresponds to **positive arousal** and generally **negative valence**.
- The **minor scale** contains intervals which are greater than half and whole steps, this corresponds to **negative valence and dominance**.

These assumptions are not meant to be comprehensive. After consulting literature, there is not a widely agreed upon method of mapping from these harmonic qualities to scales such as PAD. This is simply a starting point for our experiment; the goal is to see how participants rate the sounds in combination with animations for this particular use case (StoryScape).

Lerdahl (2001) Tonal Pitch Space. New York, NY; Oxford

# Background: Emotion words and models

But which emotions should we actually use to build our emotive shapes and sounds?

After searching the existing StoryScape stories to find commonly used emotion words, and referred to OCC model (Ortony, Clore & Collins) to find emotions that correlate with common structure of stories, we settled on the following 5 emotion words: **Calm, Happy, Angry, Sad, Sleepy.** We then used the PAD model to map these emotions from the OCC onto the valence, arousal and dominance scales that our hypothesis for emotive shape and sounds build upon (Kasap).

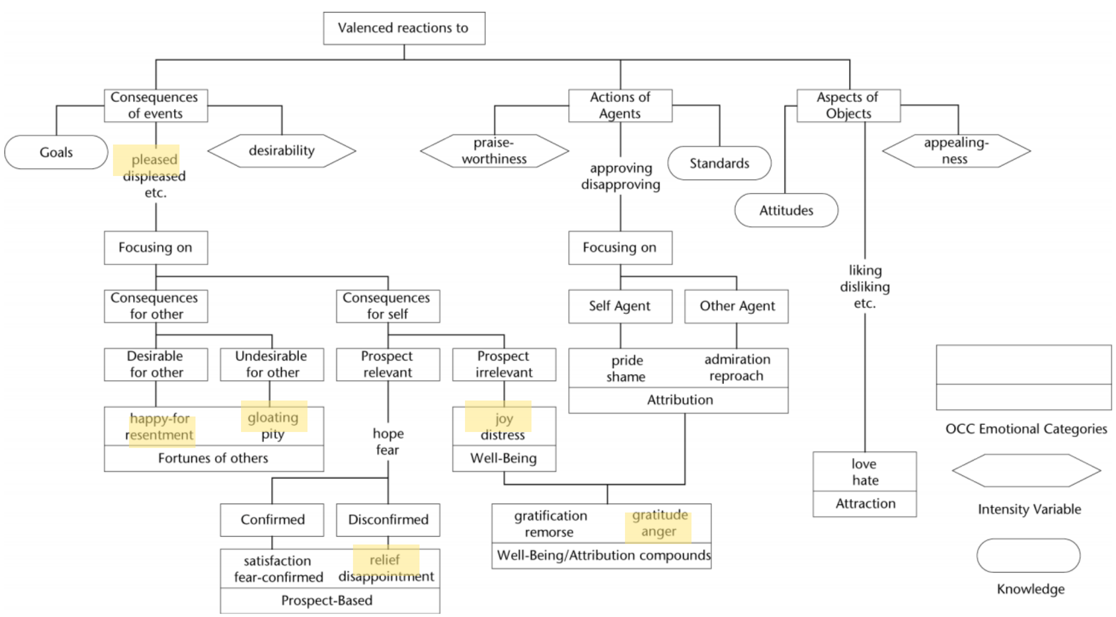



Figure 1: The original OCC model.


A. Ortony, G.L. Clore, and A. Collins, The Cognitive Structure of Emotions, Cambridge Univ. Press, 1988.


**Table 3. Mapping from OCC* emotions to PAD space.**

| Emotion | Pleasure | Arousal | Dominance | Mood type |
|---|---|---|---|---|
| Joy | 0.40 | 0.20 | 0.10 | +P+A+D Exuberant |
| Hope | 0.20 | 0.20 | –0.10 | +P+A–D Dependent |
| Relief | 0.20 | –0.30 | 0.40 | +P–A+D Relaxed |
| Pride | 0.40 | 0.30 | 0.30 | +P+A+D Exuberant |
| Gratitude | 0.40 | 0.20 | –0.30 | +P+A–D Dependent |
| Love | 0.30 | 0.10 | 0.20 | +P+A+D Exuberant |
| Happy–for | 0.40 | 0.20 | 0.20 | +P+A+D Exuberant |
| Gloating | 0.30 | –0.30 | –0.10 | +P–A–D Docile |
| Distress | –0.40 | –0.20 | –0.50 | –P–A–D Bored |
| Fear | –0.64 | 0.60 | –0.43 | –P+A–D Anxious |
| Disappointment | –0.30 | 0.10 | –0.40 | –P+A–D Anxious |
| Remorse | –0.30 | 0.10 | –0.60 | –P+A–D Anxious |
| Anger | –0.51 | 0.59 | 0.25 | –P+A+D Hostile |
| Hate | –0.60 | 0.60 | 0.30 | –P+A+D Hostile |
| Sorry–for | –0.40 | –0.20 | –0.50 | –P–A–D Bored |
| Resentment | –0.20 | –0.30 | –0.20 | –P–A–D Bored |

**Ortony, Clore, and Collins model*

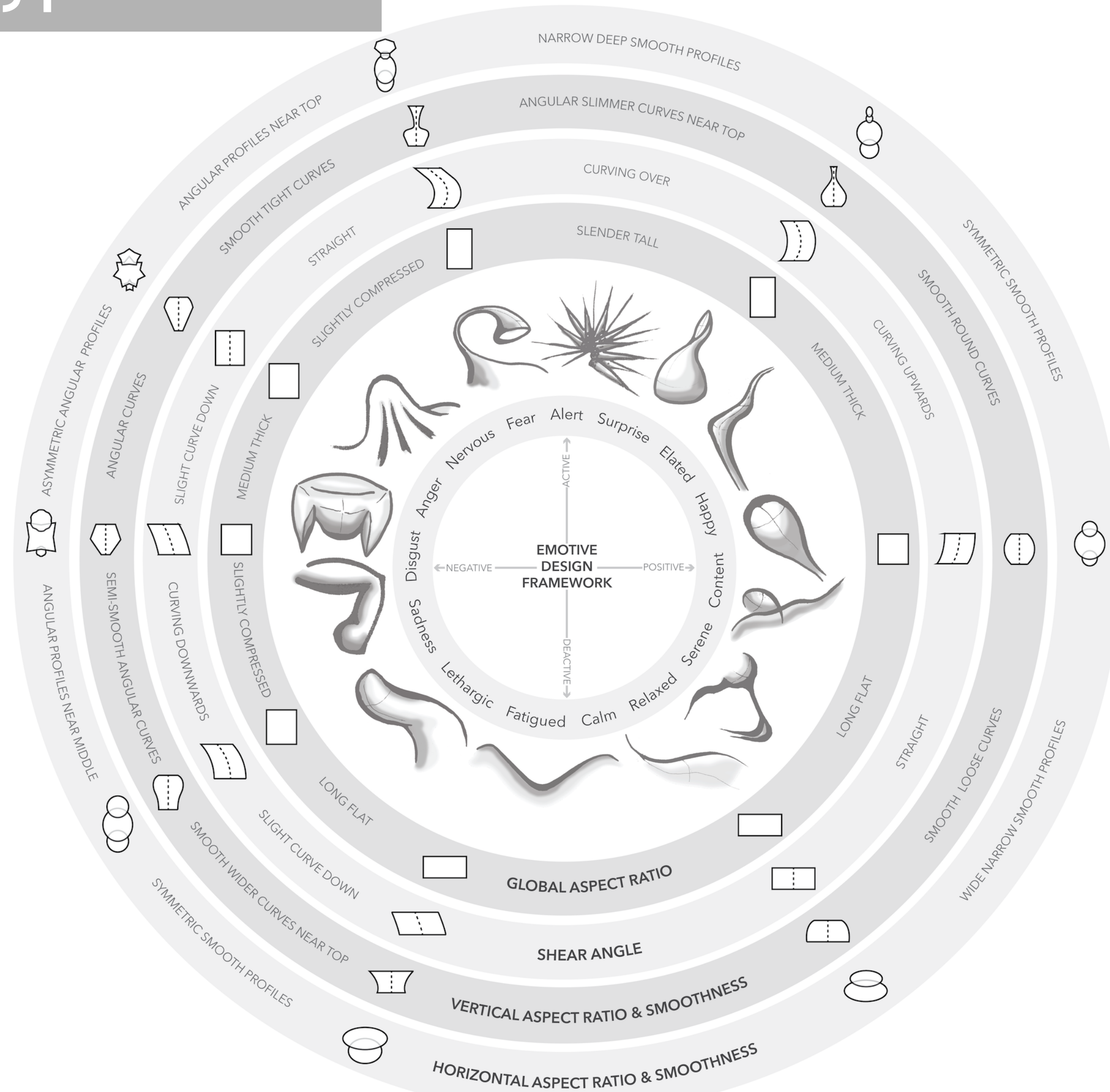

Kasap, Z., Ben Moussa, M., Chaudhuri, P., & Magnenat-Thalmann, N. (2009). Making them remember— emotional virtual characters with memory. *Computer Graphics and Applications, IEEE*, *29*(2), 20-29.

# Experimental Materials

# Emotive Shapes: Hypothesis

A framework for mapping different shape features to the PAD scales and various emotions built upon the existing research described above, i.e. smooth-angular related to positive-negative valence, tall-flat related to high-low arousal, leaning forward-backward related to positive-negative dominance.

This framework breaks these shapes down into detailed design elements such as aspect ratio, smoothness, shear etc. which can then be used to transform a single shape into a range of emotive variants of that shape.





Collier, G. L. (1996). Affective synesthesia: Extracting emotion space from simple perceptual stimuli. *Motivation and Emotion,* 20(1), 1-32.

Isbister, K. et al. (2007). The sensual evaluation instrument: Developing a trans-cultural self-report measure of affect *International journal of human-computer studies*, 65(4)

Poffenberger, A. T., & Barrows, B. E. (1924). The Feeling Value of Lines. *Journal of Applied Psychology*, 8(2), 187.

Russell, James A. A circumplex model of affect. *Journal of personality and social psychology* 39.6 (1980): 1161-1178

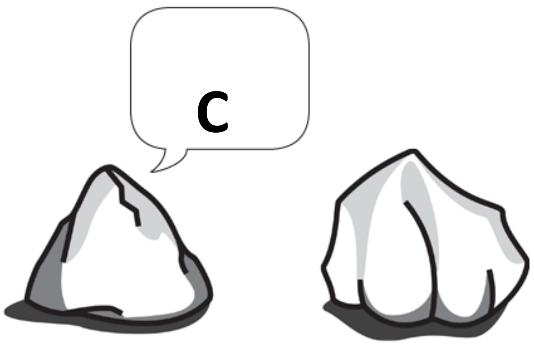
# Emotive Shapes: Hypothesis

Our hypothesis for the PAD values and related shape features for each of the 5 emotive shapes we included in our experiment are shown below:

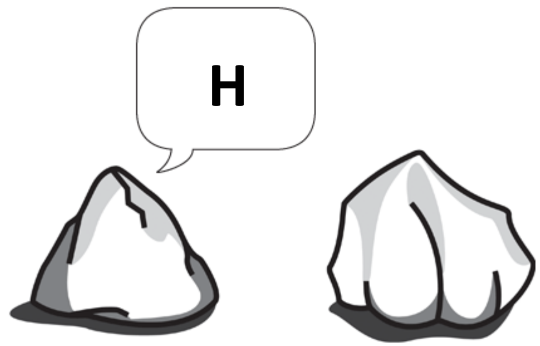
| SHAPE-EMOTION HYPOTHESIS | | | | | | | |
|---|---|---|---|---|---|---|---|
| | PAD scales | | | Shape features | | | |
| **Emotion** | **Pleasure** | **Arousal** | **Dominance** | **Smoothness** | **Aspect ratio** | **Direction** | **Animation** |
| Calm | 1 | -1 | -1 | Smooth rounded | Flat (High AR) | Forwards | Slow to fast |
| Happy | 1 | -1 | -1 | Rounded | Tall (Low AR) | Forwards | Fast to slow |
| Angry | 1 | 1 | 1 | Angular | Tall (Low AR) | Forwards | Slow to fast |
| Sad | -1 | 1 | 1 | Rounded | Flat (High AR) | Backwards | Slow to fast |
| Sleepy | -1 | -1 | -1 | Smooth rounded | Flat (High AR) | Backwards | Fast to slow |

# Emotive Shapes: Stimuli

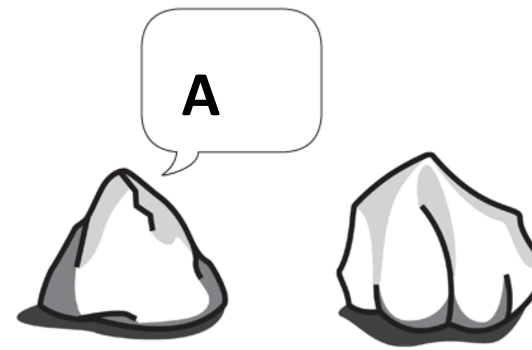
Below are the 5 emotive speech bubbles we created as stimuli for our experiment.
We attached the speech bubble to an illustration of a rock conversing with another rock to lend the context of a dialogue between two objects without implicit emotional content.



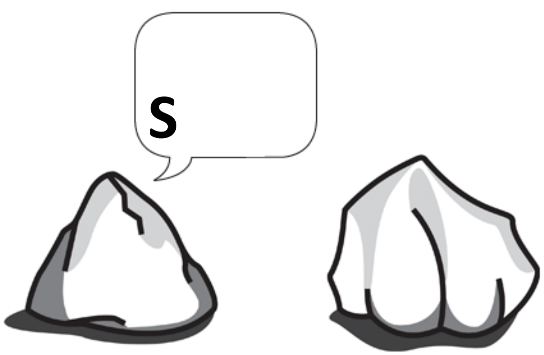

**Calm (C)**
PAD: + valence, - arousal, + dominance
Shapes: rounded, flat, leaning forward



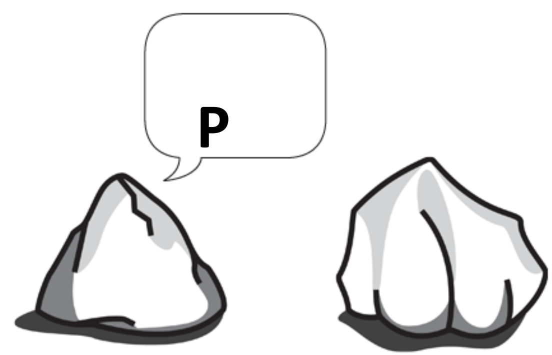

**Happy (H)**
PAD: + valence, + arousal, + dominance
Shapes: rounded, tall, leaning forward




**Angry (A)**
PAD: - valence, + arousal, + dominance
Shapes: angular, tall, leaning forward




**Sad (S)**
PAD: - valence, - arousal, - dominance
Shapes: rounded, flat, leaning backward




**Sleepy (P)**
PAD: + valence, - arousal, - dominance
Shapes: rounded, flat, leaning backward

# Emotive Sounds: Hypothesis

We created a hypothesis limited in scope by using four familiar scales (Diatonic Major, Minor, Heptatonic, Whole Tone) and synchronizing them to frames of the animation. These scales could be played up or down, and the timing of animation and scale notes were locked together. Our hypothesis for the PAD values and related emotion for each of the scale combinations we included in our experiment are shown below:

| SOUND-EMOTION HYPOTHESIS | | | | | |
|---|---|---|---|---|---|
| **Scale** | **Scale_direction** | **Pleasure** | **Arousal** | **Dominance** | **Related Emotion** |
| diatonic | up | 1 | 1 | 1 | Happy |
| diatonic | down | 1 | -1 | -1 | Sleepy/Calm |
| heptatonic | up | 1 | 1 | 1 | Happy/Sleepy |
| heptatonic | down | -1 | 1 | 1 | Angry/Sleepy |
| minor | up | -1 | -1 | -1 | Sad |
| minor | down | -1 | -1 | -1 | Sad |
| whole_tone | up | -1 | 1 | 1 | Angry |
| whole_tone | down | -1 | 1 | -1 | Frightened/Angry |


Trochidis K., Delbé C., & Bigand E (2011) INVESTIGATION OF THE RELATIONSHIPS BETWEEN AUDIO FEATURES AND INDUCED EMOTIONS IN CONTEMPORARY WESTERN MUSIC, SMC Conference

Madsen J (2011) Modeling of Emotions expressed in Music using Audio features, Informatics and Mathematical Modelling, Technical University of Denmark

Laurier, C., Lartillot O., Eerola T., & Toiviainen P. (2009) Exploring Relationships between Audio Features and Emotion in Music, ESCOM, Conference of European Society for the Cognitive Sciences of Music

Sezgin et al. (2012) Perceptual audio features for emotion detection, EURASIP Journal on Audio, Speech, and Music Processing 2012:16

# Emotive Sounds: Stimuli

All sound and scales were generated programmatically in the web browser, and so could easily be integrated into the StoryScape app.

**Generating sounds in the web browser**

Sample generation is achieved in the browser using Keith Horwood's open source Audio Synth ( https://github.com/keithwhor/audiosynth). The artificially synthesized piano sample for each note is intended to be a relatively neutral sound. It does not have a particularly long or short attack or decay and it is relatively pure-toned, not dissonant or abrasive. It is intended not to affect perception of the stability of the scale being played. The timing, direction, and scale type should have much stronger influence. The generated sound was the same for all participants and does not vary based on browser implementation.

Line 225 defines attack, decay, sustain and noise functions that are used to play each note of the scale.

Each of the emotive shapes was also left silent in each study so as to gauge the emotional perception of the shapes alone, and understand the effect of any sounds on that perception.

# Experimental Methods

# User Evaluation: Study Design

Our study was carried out by creating a website with a survey asking participants to evaluate the different shape and sound stimuli. We had originally intended on distributing a link to the website to many participants via email, and having participants contribute any time or place. Instead, we ran individual sessions with participants in controlled environments to limit the level of unwarranted distractions that could occur.

Using pre-questions, we attempted to ascertain the participant's self reported mood using the Brief Mood Introspection Scale (Mayer). This mood rating scale is a freely available model consisting of 16 mood-adjectives (e.g. "happy", "tired", "caring") to which a person rates how they are feeling on a 4 point scale ("Definitely feel", "Slightly feel", "Do not feel", "Definitely do not feel"). Responses to the scale can then yield measures of overall pleasant-unpleasant mood and arousal-calm mood of each of the participants.

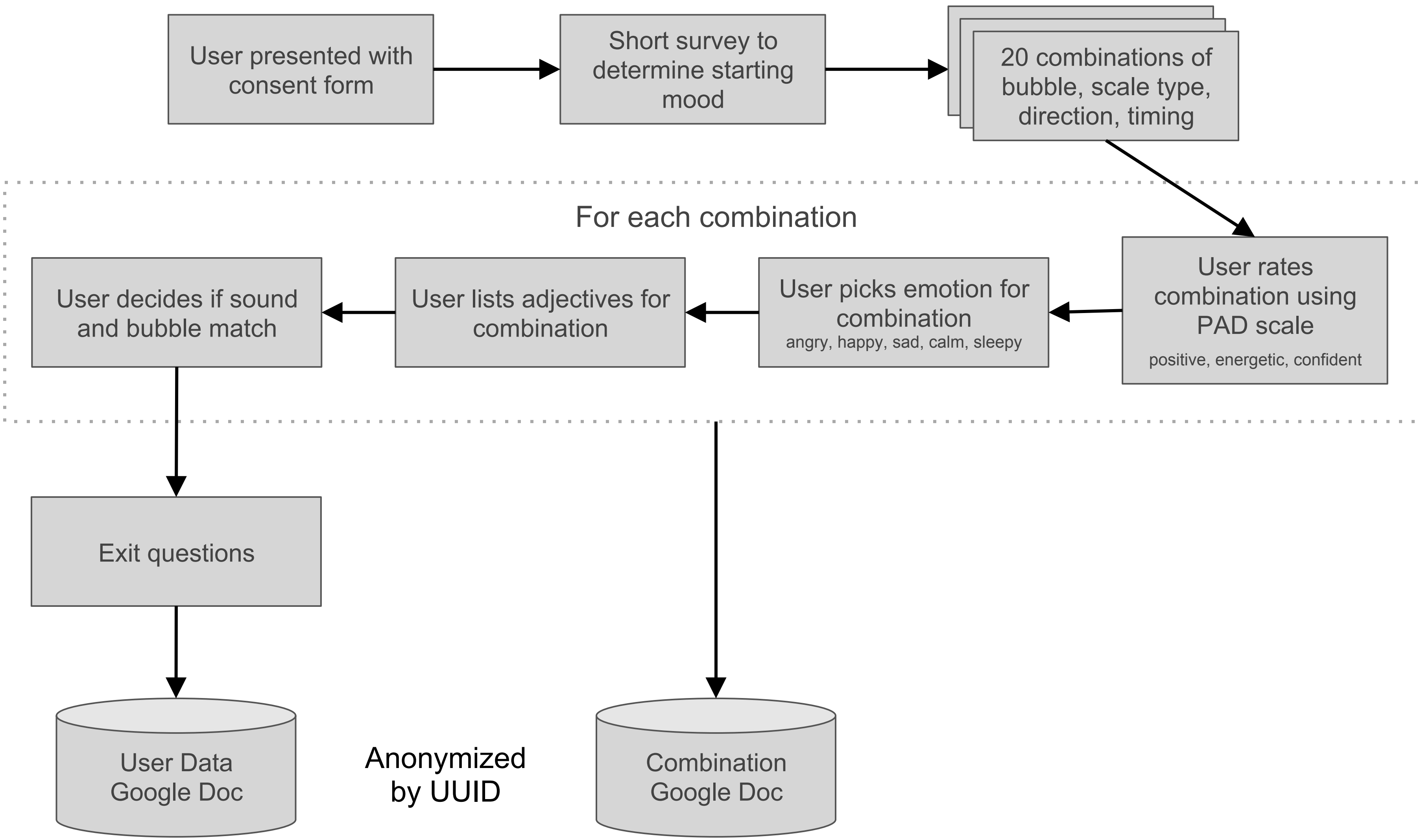
The participants then continued to the main evaluation where they were asked to view different combinations of shape and sound stimuli. They were then asked to rate how they perceived certain emotional attributes of the speech bubble. The three PAD scales were tested by asking users to rate the "Positivity", "Energy" and "Confidence" of the speech bubbles on a 5 point scale (Very…, Slightly…, Neutral, Slightly Un…, Very Un…). The participants were also asked to chose one of the 5 emotions (Calm, Happy, Angry, Sad, Sleepy) which best represented the emotion of the shape-sound animation – this was asked after the PAD scales so as not to bias the responses to these questions. Finally, the participants were asked to fill in any other adjectives they thought suited the animation, and rate if they thought the sound matched the shape. After the participants had evaluated all 20 shape-sound combinations, they were asked two questions regarding what they thought of the idea of using emotive shapes and sounds to communicate emotions in a storytelling context.

The study took a total of 15-20 minutes per participant. 25 participants took part in the study, leading to us collecting 500 individual responses to the shape-sound stimuli. The anonymized data collected from the website survey was collected into two Google docs for further analysis.

Mayer, J. D., & Gaschke, Y. N. (1988). The experience and meta-experience of mood. Journal of Personality and Social Psychology, 55, 102-111.

# User Evaluation: Study Design

Here is flow chart describing the protocol we used in our study:

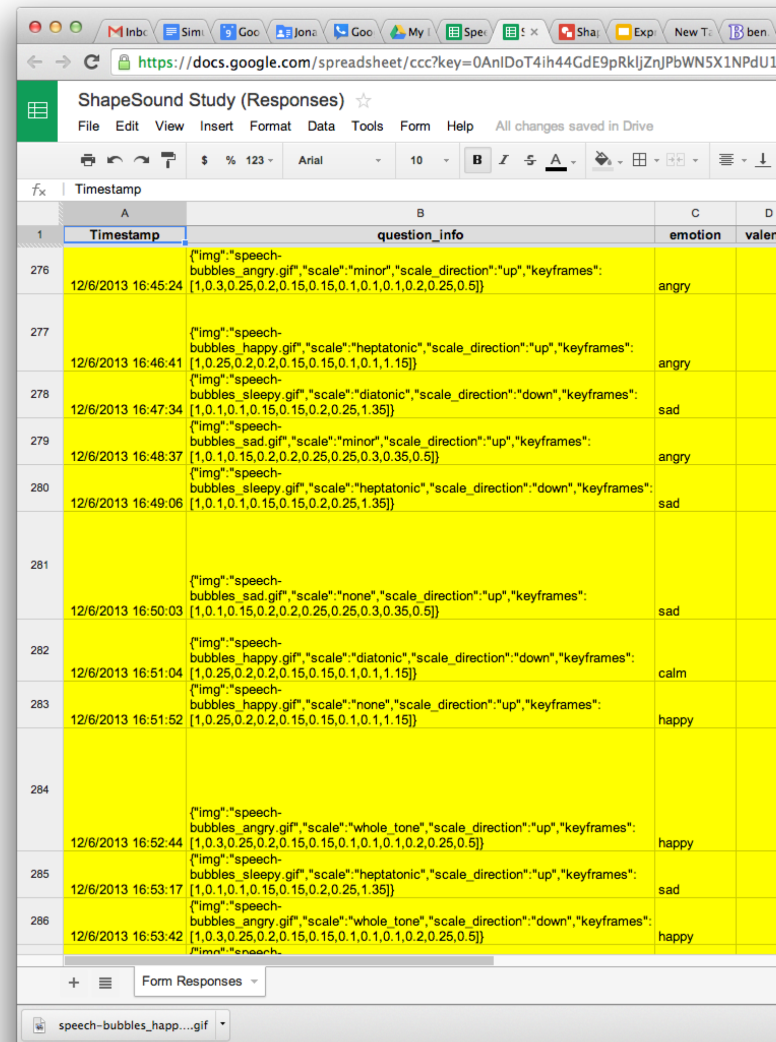


# User Evaluation: Website Design

In the website, we programmatically define (in JSON) a combination of animation shape, timing, sound, scale type and scale direction. User responses are direct outputted to google docs.

Whilst random combinations can be generated, we created 20 intentional combinations to represent a good cross section of the scales and bubbles. These combinations were presented in random order to reduce biasing the responses due to emotional priming.

```
{"questions":[
          {
                    "img":"speech-bubbles_calm",
                    "scale":"diatonic",
                    "scale_direction":"up",
                    "keyframes":[0.35,0.3,0.2,0.15,0.1,0.1,2]
          },
          {
                    "img":"speech-bubbles_happy",
                    "scale":"whole_tone",
                    "scale_direction":"up",
                    "keyframes":[0.25,0.2,0.2,0.15,0.15,0.1,0.1]
          }
],
          "scales": {
                    "diatonic": [ "C", "D", "E", "F", "G", "A", "B"],
                    "whole_tone": [ "C", "D", "E", "F#", "G#", "A#"],
                    "minor" : ["C", "D", "D#", "F", "G", "G#", "B"],
                    "heptatonic": ["C", "C#", "E", "F", "G",
"G#", "B"]
          }
}
```



ShapeSound Study (Responses)

| | A | B | C | D |
|---|---|---|---|---|
| 1 | Timestamp | question_info | emotion | valen |
| 276 | 12/6/2013 16:45:24 | {"img":"speech-bubbles_angry.gif","scale":"minor","scale_direction":"up","keyframes":[1,0.3,0.25,0.2,0.15,0.15,0.1,0.1,0.1,0.2,0.25,0.5]} | angry | |
| 277 | 12/6/2013 16:46:41 | {"img":"speech-bubbles_happy.gif","scale":"heptatonic","scale_direction":"up","keyframes":[1,0.25,0.2,0.2,0.15,0.15,0.1,0.1,1.15]} | angry | |
| 278 | 12/6/2013 16:47:34 | {"img":"speech-bubbles_sleepy.gif","scale":"diatonic","scale_direction":"down","keyframes":[1,0.1,0.1,0.15,0.15,0.2,0.25,1.35]} | sad | |
| 279 | 12/6/2013 16:48:37 | {"img":"speech-bubbles_sad.gif","scale":"minor","scale_direction":"up","keyframes":[1,0.1,0.15,0.2,0.2,0.25,0.25,0.3,0.35,0.5]} | angry | |
| 280 | 12/6/2013 16:49:06 | {"img":"speech-bubbles_sleepy.gif","scale":"heptatonic","scale_direction":"down","keyframes":[1,0.1,0.1,0.15,0.15,0.2,0.25,1.35]} | sad | |
| 281 | 12/6/2013 16:50:03 | {"img":"speech-bubbles_sad.gif","scale":"none","scale_direction":"up","keyframes":[1,0.1,0.15,0.2,0.2,0.25,0.25,0.3,0.35,0.5]} | sad | |
| 282 | 12/6/2013 16:51:04 | {"img":"speech-bubbles_happy.gif","scale":"diatonic","scale_direction":"down","keyframes":[1,0.25,0.2,0.2,0.15,0.15,0.1,0.1,1.15]} | calm | |
| 283 | 12/6/2013 16:51:52 | {"img":"speech-bubbles_happy.gif","scale":"none","scale_direction":"up","keyframes":[1,0.25,0.2,0.2,0.15,0.15,0.1,0.1,1.15]} | happy | |
| 284 | 12/6/2013 16:52:44 | {"img":"speech-bubbles_angry.gif","scale":"whole_tone","scale_direction":"up","keyframes":[1,0.3,0.25,0.2,0.15,0.15,0.1,0.1,0.1,0.2,0.25,0.5]} | happy | |
| 285 | 12/6/2013 16:53:17 | {"img":"speech-bubbles_sleepy.gif","scale":"heptatonic","scale_direction":"up","keyframes":[1,0.1,0.1,0.15,0.15,0.2,0.25,1.35]} | sad | |
| 286 | 12/6/2013 16:53:42 | {"img":"speech-bubbles_angry.gif","scale":"whole_tone","scale_direction":"down","keyframes":[1,0.3,0.25,0.2,0.15,0.15,0.1,0.1,0.1,0.2,0.25,0.5]} | happy | |

Form Responses

# User Evaluation: Website Design

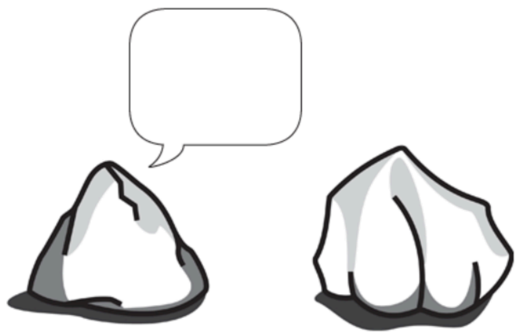
Here is an example of one of the animation evaluation pages of the website:

Click here to try it for yourself:
[http://ben.ai/mas630/scale.html?data=shapes_sounds_01.json](http://ben.ai/mas630/scale.html?data=shapes_sounds_01.json)

Hey, thanks for helping us! Please put on your headphones and click "Run Animation" below.

**Animation 1/20**

Run animation



Select how positive you think the emotion that the speech bubble is expressing is:

| Very Positive | Slightly Positive | Neutral | Slightly Negative | Very Negative |
|---|---|---|---|---|

Select how energetic you think the emotion that the speech bubble is expressing is:

| Very Energetic | Slightly Energetic | Neutral | Slightly Unenergetic | Very Unenergetic |
|---|---|---|---|---|

Select how confident you think the emotion that the speech bubble is expressing is:

| Very Confident | Slightly Confident | Neutral | Slightly Unconfident | Very Unconfident |
|---|---|---|---|---|

Select which emotion you think the speech bubble is expressing:

| Angry | Calm | Happy | Sad | Sleepy |
|---|---|---|---|---|

Please suggest any other adjectives which could describe this speech bubble:

If there was sound with the animation, did you think the music matched the shapes?

| Yes | Somewhat | No |
|---|---|---|

# Results

# Results: Overview by shape

Here is a summary of how each of the shape-sound combinations were rated for overall emotion. Values in red show very low numbers of responses for a certain shape-sound combination, and values in green show very high numbers of responses. (Table continued on next page…)

| ***Bubble*** | ***Scale*** | ***Calm*** | ***Happy*** | ***Angry*** | ***Sad*** | ***Sleepy*** |
|---|---|---|---|---|---|---|
| A | Minor, Up | **0%** | 48% | 52% | **0%** | **0%** |
| A | Silent | **4%** | 48% | 48% | **0%** | **0%** |
| A | Whole Tone, Down | **8%** | **8%** | 52% | 24% | **8%** |
| A | Whole Tone, Up | **0%** | 68% | 32% | **0%** | **0%** |
| ***A Total*** | | ***3%*** | *43%* | *46%* | ***6%*** | ***2%*** |
| C | Diatonic, Down | **8%** | **0%** | **0%** | 72% | 20% |
| C | Diatonic, Up | 20% | 20% | **4%** | 32% | 24% |
| C | Minor, Down | **0%** | **0%** | **0%** | 76% | 24% |
| C | Silent | **12%** | **8%** | **0%** | 52% | 28% |
| ***C Total*** | | ***10%*** | ***7%*** | ***1%*** | *58%* | *24%* |
| H | Diatonic, Down | 36% | 36% | **4%** | 20% | **4%** |
| H | Diatonic, Up | **12%** | 84% | **4%** | **0%** | **0%** |
| H | Heptatonic, Up | **16%** | 72% | **8%** | **4%** | **0%** |
| H | Silent | 44% | 56% | **0%** | **0%** | **0%** |
| ***H Total*** | | *27%* | *62%* | ***4%*** | ***6%*** | ***1%*** |

# Results: Overview by shape

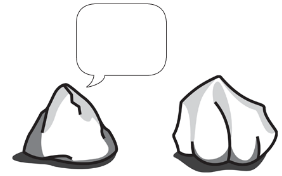
| Bubble | Scale | Calm | Happy | Angry | Sad | Sleepy |
|---|---|---|---|---|---|---|
| S | Diatonic, Down | 28% | **4%** | **16%** | 44% | **8%** |
| S | Minor, Down | **12%** | **0%** | **12%** | 64% | **12%** |
| S | Minor, Up | **16%** | 24% | 20% | 32% | **8%** |
| S | Silent | 36% | **0%** | **8%** | 36% | 28% |
| ***S Total*** | | *23%* | ***7%*** | ***14%*** | *44%* | ***12%*** |
| P | Diatonic,Down | **12%** | **0%** | **0%** | **76%** | **12%** |
| P | Heptatonic,Down | **4%** | **0%** | **4%** | 68% | 20% |
| P | Heptatonic, Up | 20% | **0%** | **12%** | 48% | 20% |
| P | Silent | 20% | **0%** | **0%** | 44% | 36% |
| ***P Total*** | | ***14%*** | ***0%*** | ***4%*** | *59%* | *22%* |

This seems to show that our hypothesis were not very effective at communicating anything but the most basic emotions. It is very easy to understand from this chart that there are several combinations which are clearly not effective at conveying certain emotions. There are a few combinations which are effective at conveying happy and sad emotions. In the following slides, we'll take a deeper look and break down these perceptions by PAD to understand if components of the combinations were expressive in isolation.

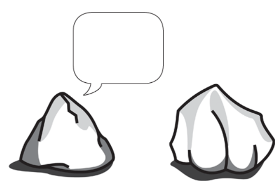
# Results: Calm

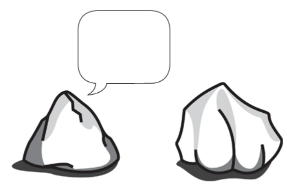
Calm hypothesis:

- Shape: +P, -A, -D (rounded, flat, forward)
- Sound: diatonic down



- Percentage of all combinations rated as calm: 15.4% (lower than the 20% expected)
- Very few participants perceived the C bubble as Calm
- Calm was perceived most for the H (31.5%) and S (29.9%) bubbles



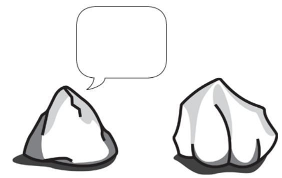


- Calm was perceived for diatonic down scale (20.8%)

# Results: Calm

Calm hypothesis:

- Shape: +P, -A, -D (rounded, flat, forward)
- Sound: diatonic down



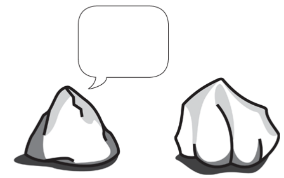
There was generally poor perception of the C bubble as its hypothetical emotion Calm. Both with and without sound, the animations were rated mainly as Sad, leading to a very low match to our hypotheses. Our hypothesized Valence and Dominance ratings of the animations (with and without sound) were also not perceived highly. The Valence of the animation was in-fact perceived as opposite to our hypothesis (negative instead of positive). However, the negative Arousal of the animation was perceived correctly and by a large proportion of the participants (over 76% for both with and without sound).

| Perception of C bubble | Perceived emotion (avg) | Emotion perception match hypothesis? | Perceived valence (avg) | Perceived arousal (avg) | Perceived dominance (avg) |
|---|---|---|---|---|---|
| **No sound** | Sad (52%) | 16% | -1 (8% c/w hypothesis) | -1 (76% c/w hypothesis) | -1 (12% c/w hypothesis) |
| **With sound: diatonic down** | Sad (72%) | 8% | -1 (0% c/w hypothesis) | -1 (84% c/w hypothesis) | -1 (8% c/w hypothesis) |

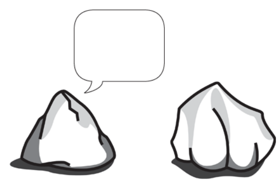
# Results: Happy

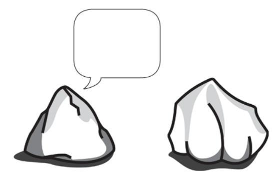
Happy hypothesis:

- Shape: +P, +A, +D (rounded, tall, forward)
- Sound: diatonic up, heptatonic up



- Percentage of all combinations rated as happy: 23.6% (higher than the 20% expected)
- Over half of participants perceived the H bubble as Happy
- Happy perceived most for the H (52.5%) and A (36.4%) bubbles



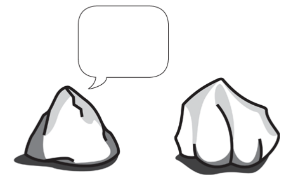


- Happy perceived for diatonic up (17.8%), heptatonic up (15.3%)
  (A bubbles with wholetone up and minor up scales were also perceived by some participants as happy)

# Results: Happy

Happy hypothesis:

- Shape: +P, +A, +D (rounded, tall, forward)
- Sound: diatonic up, heptatonic up



Our hypotheses for Happy shapes and sounds performed much better. All of our hypotheses were perceived correctly for over half the participants in all of the categories. The Valence perceived especially correlated with our hypotheses, with over 76% of participants agreeing with our hypotheses.

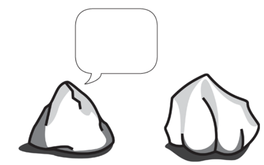
With sound, there was a general increase in participants agreeing with our hypotheses. The **diatonic up scale** especially seemed to enhance the Happy perception by the participants (all scores above 72%).

| Perception of H bubble | Perceived emotion (avg) | Emotion perception match hypothesis? | Perceived valence (avg) | Perceived arousal (avg) | Perceived dominance (avg) |
|---|---|---|---|---|---|
| **No sound** | Happy (56%) | 56% | 1 (80% c/w hypothesis) | 1 (60% c/w hypothesis) | 1 (52% c/w hypothesis) |
| **With sound: diatonic up** | Happy (84%) | 84% | 1 (92% c/w hypothesis) | 1 (88% c/w hypothesis) | 1 (72% c/w hypothesis) |
| **With sound: heptatonic up** | Happy (72%) | 72% | 1 (76% c/w hypothesis) | 1 (80% c/w hypothesis) | 1 (60% c/w hypothesis) |

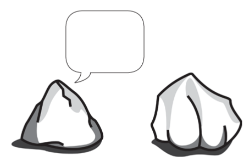
# Results: Angry

Angry hypothesis:

- Shape:  -P, +A, +D (angular, tall, forward)
- Sound:  wholetone up & down (heptatonic down not tested)



- Percentage of all combinations rated as angry: 13.8% (lower than the 20% expected)
- Over two-thirds of participants perceived the A bubble as Angry
- Angry perceived most for the A (68.1%) bubble

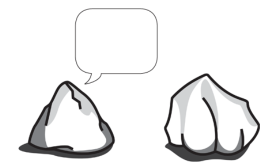


- Angry perceived for wholetone down (18.8%), wholetone up (11.6%)
  (The minor up scale – what we hypothesis to be a Sad sound - was also considered to be Angry when paired with the A bubble (18.8%))

# Results: Angry

Angry hypothesis:

- Shape: -P, +A, +D (angular, tall, forward)
- Sound: wholetone up & down (heptatonic down not tested)



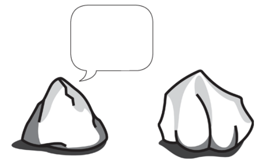
Our hypotheses for Angry shapes and sounds had a mixed performance. Our hypotheses for pure emotion were only perceived correctly for around half the participants. When there was no sound, almost half of the participants thought the shape looked Angry and almost half thought it looked Happy. In general, our hypothesized valence was perceived poorly – people generally thought the shape had a positive valence instead of our negative valence hypothesis. However, over 80% of participants agreed with our Arousal and Dominance hypotheses for most of the animations (except with the wholetone down scale). The **minor up scales** seemed to enhance the Angry perception by the participants. However, the **wholetone up and wholetone down scale** did not create as strong correlations to the Angry shape – wholetone up was in-fact considered Happy rather than Angry.

| Perception of A bubble | Perceived emotion (avg) | Emotion perception match hypothesis? | Perceived valence (avg) | Perceived arousal (avg) | Perceived dominance (avg) |
|---|---|---|---|---|---|
| **No sound** | Happy (48%)<br>Angry (48%) | 48% | 1<br>(32% c/w hypothesis) | 1<br>(84% c/w hypothesis) | 1<br>(84% c/w hypothesis) |
| **With sound: wholetone up** | Happy (68%) | 32% | 1<br>(28% c/w hypothesis) | 1<br>(100% c/w hypothesis) | 1<br>(92% c/w hypothesis) |
| **With sound: wholetone down** | Angry (52%) | 52% | -1<br>(60% c/w hypothesis) | 1<br>(60% c/w hypothesis) | 1<br>(48% c/w hypothesis) |
| **With sound: minor up** | Angry (52%) | 52% | 1<br>(28% c/w hypothesis) | 1<br>(88% c/w hypothesis) | 1<br>(84% c/w hypothesis) |

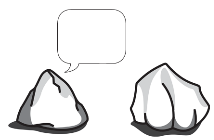
# Results: Sad

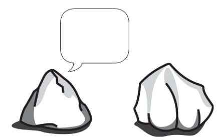
Sad hypothesis:

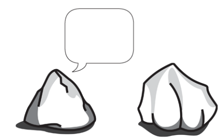
- Shape: -P, -A, -D (rounded, flat, backwards)
- Sound: minor up & down



- Percentage of all combinations rated as sad: 34.9% (higher than the 20% expected)
- Only a third of participants perceived the S bubble as Sad
- Sad perceived most for the S (25.3%), C (33.3%) and P (34.5%) bubbles





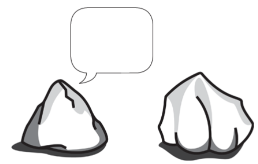


- Sad somewhat perceived for minor down (9.2%) and minor up (4.6%)
(calm bubbles with minor down or diatonic down, and sleepy bubbles with diatonic down and heptatonic down sounds were also perceived as sad)

# Results: Sad

Sad hypothesis:

- Shape: -P, -A, -D (rounded, flat, backwards)
- Sound: minor up & down



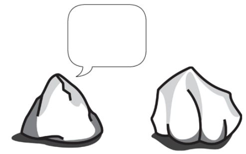
Our hypotheses for Sad shapes and sounds did not perform well. Our hypotheses for pure emotion were perceived correctly less than a third of the participants. When there was no sound, a third of the participants perceived the shape to be Sad, and another third perceived it to be Calm. In general, our hypothesized Valence and Dominance ratings were perceived fairly well. However, less than half of the participants agreed with our Arousal hypotheses. The **minor down scales** seemed to enhance the Sad perception by the participants, especially the Valence rating. However, the **minor up scale** did not create as strong correlations to the Sad shape – the participants in-fact considered the Arousal rating for this combination to be positive instead of negative.

| Perception of S bubble | Perceived emotion (avg) | Emotion perception match hypothesis? | Perceived valence (avg) | Perceived arousal (avg) | Perceived dominance (avg) |
|---|---|---|---|---|---|
| **No sound** | Sad (36%)<br>Calm (36%) | 36% | -1<br>(60% c/w hypothesis) | -1<br>(44% c/w hypothesis) | -1<br>(80% c/w hypothesis) |
| **With sound: minor up** | Sad (32%) | 32% | -1<br>(48% c/w hypothesis) | 1<br>(24% c/w hypothesis) | -1<br>(52% c/w hypothesis) |
| **With sound: minor down** | Sad (64%) | 64% | -1<br>(84% c/w hypothesis) | -1<br>(48% c/w hypothesis) | -1<br>(72% c/w hypothesis) |

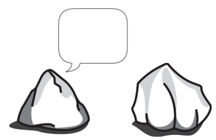
# Results: Sleepy

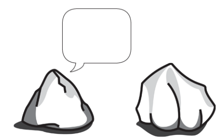
Sleepy hypothesis:

- Shape: P, -A, -D (rounded, flat, backwards)
- Sound: diatonic down, heptatonic up & down



- Percentage of all combinations rated as sleepy: 12.2% (higher than the 20% expected)
- Very few participants perceived the P bubble as Sleepy
- Sad perceived for the P 36.1% and C 39.3% bubbles



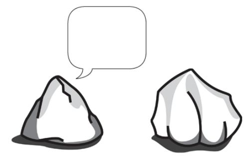


- Sad somewhat perceived for heptatonic up (8.2%), heptatonic down (8.2%), diatonic down (4.9%) (calm bubbles with diatonic down or minor down were also perceived as sad)

# Results: Sleepy

Sleepy hypothesis:

- Shape: P, -A, -D (rounded, flat, backwards)
- Sound: diatonic down, heptatonic up & down



There was generally poor perception of the P bubble as its hypothetical emotion Sleepy. Both with and without sound, the animations were rated mainly as Sad, leading to a very low match to our hypotheses. Our hypothesized Valence ratings of the animations (with and without sound) were also not perceived highly – most participants in-fact rated the valence as opposite to our hypothesis (negative rather than positive). However, the negative Arousal and Dominance of the animation was perceived correctly and by a large proportion of the participants for most of the animations (over 76% for both with and without sound). Whilst none of the sounds appeared to greatly enhance the perception of our perceived emotions, the hepatonic up scale seemed to confuse participants as many fewer correctly perceived our hypothesized Arousal and Dominance ratings.

| Perception of P bubble | Perceived emotion (avg) | Emotion perception match hypothesis? | Perceived valence (avg) | Perceived arousal (avg) | Perceived dominance (avg) |
|---|---|---|---|---|---|
| **No sound** | Sad (44%) | 36% | -1 (8% c/w hypothesis) | -1 (80% c/w hypothesis) | -1 (76% c/w hypothesis) |
| **With sound: diatonic down** | Sad (76%) | 12% | -1 (4% c/w hypothesis) | -1 (80% c/w hypothesis) | -1 (76% c/w hypothesis) |
| **With sound: heptatonic up** | Sad (48%) | 20% | -1 (4% c/w hypothesis) | -1 (52% c/w hypothesis) | -1 (52% c/w hypothesis) |
| **With sound: heptatonic down** | Sad (68%) | 20% | -1 (0% c/w hypothesis) | -1 (96% c/w hypothesis) | -1 (80% c/w hypothesis) |

# Conclusions

# Results

**In general:**

- Perception of arousal and dominance matched the hypothesis much more than valence. This agrees with a lot of other research; it is relatively simple to convey energy and confidence, but much more difficult to communicate whether the energy and confidence are positive or negative. These cues for valence often come from context.

- The correlation of participant's perceptions to the hypotheses improved when the hypotheses for sound and shape were similar on the PAD scale.

**Emotive shapes:**

- Happy, Angry and Sad shapes seem to correlate reasonably well with the hypotheses. There were several combinations of shapes for each that clearly communicated those emotions for a high percentage of participants.

- Calm and Sleepy seem to not correlate well with the hypotheses. It possible these emotions are too nuanced to be communicated effectively with this type of test. The majority of the combinations tried were not perceived as calm or happy.

**Emotive sounds:**

- Adding sound resulted in an amplification of the majority perceived emotion in all but one combination, regardless of scale and direction or whether the perceived emotion matched any hypothesis. In this sense, it is clear that sound provides a benefit to expression in this medium.

# Results

There was quite a strong response to our exit survey questions. Because our participants were Media Lab students, it is probable that the group was heavily biased, but participants seemed to enjoy categorizing the shapes even after 20 repeating sounds and forms.

- In general, do you think the animated and audio speech bubbles could enhance how dialogue in a visual illustration could convey emotion?

  **Yes: 84%** **Somewhat: 16%** **No:0 %**

- Do you think this sort of expressive animation could help you to convey more emotion in an online illustrated story you might be writing?

  **Yes: 84%** **Somewhat: 12%** **No: 4%**

We could conclude that including animated expressive speech bubbles like this in an app such as StoryScape could help people to convey more emotion in their stories.

# General Learning & Limitations

**Study design:**

- A controlled environment ensures our results are much better normalized than if we were to send the study out to the general public. Initially, the hope was to test more combinations with a greater number of people. However, we decided to contribute only a
- A few participants needed some of the words translated during the study, so any further studies must ensure that the language used is comprehensible to all participants, and offer translations if needed
- Despite a long survey at the beginning of a study possibly acting as a good mood leveler and priming people to get into an introspective mood, the self-reported mood data we collected didn't appeal to give very useful data. The hope was that using the BMIS system would provide some insight on participants initial mood. We have learned that self reporting is not necessarily the most effective way to measure emotion and this proved true. However, a long list of 16 questions was a good way to neutralize people before starting the process of rating bubbles.
- The system we created was capable of mixing and matching every aspect of the animation. With so many options for four parameters, we would have had to give users 160 combinations to exhaustively test for all possibilities. Instead we chose to test a smaller subset of the possible content. The 20 combinations we displayed were carefully chosen to represent what we hoped would be a representative set of emotions. Since we had trouble identifying the more nuanced emotions, it would be a good next step to test more combinations.
- For future versions of the study, we would like to keep track of the number of times an animation is viewed and how long it takes the user to complete each section.

**Animations:**

- If it were possible to have a completely neutral subject for the speech bubbles, that would be ideal. Our highest perceived emotion was sadness and this could have been due to the nature of the rocks.  Future studies could use stick figures.

**Sounds:**

- To find a mapping from the scales to PAD is not trivial. We made an attempt but our data does not indicate this mapping is particularly effective. If anything it proved that adding sound in any form helps the expressiveness of the animation. Scales as a musical structure were appropriately narrow in scope, and easy to test, but a full implementation in story scape might want to consider more complex melodies. These could be generatively tested in much the same way as the scales.

# Future Work

**In the future, we'd like to continue this research with the following additions:**

- If COUHES permits (may require a new proposal) we would like to distribute this tool online and get many more responses. The hope is that if enough responses can be collected, the data would be able to normalize itself with respect to context and environment.
- For a larger study, we would hope to use randomly generated combinations of timings, scales, and bubbles. This provides a very large body of tests to give users.
- Currently, timings are attached to bubbles. The system has the capability to define different timings given a single set of speech bubble animations. Taking advantage of this, the timing could be paired with bubbles in the same way as the scales.
- This system runs entirely in javascript with a single HTML file. This should be easily portable to the Storyscape engine.
- Once we have an idea of the expressiveness of various shape/scale/timing combinations they could be added to Storyscape with text entry. At this point a study could be conducted to analyze how the shapes are used together with text. Does a changing shape/scale/timing affect how text in the bubbles is perceived?
- Once the bubbles are integrated with Storyscape, some analysis could done to understand if people are using the bubbles in the way the corresponds to the models we chose.

Thank you

Further questions:

pip@mit.edu, benb@mit.edu