Mixed Company

Carey Dodge
Sonic Arts Research Centre
Queen's University Belfast
Belfast BTZ 1NN
Northern Ireland

doodgeroo@yahoo.ca

ABSTRACT

In this paper, I will describe an installation entitled, *Mixed Company*. The installation explored ideas of personal space using recorded speech, speakers and motion capture. I will describe the various approaches and techniques used both for the creation of text and the technical apparatus then discuss strengths and weakness of the resulting installation.

Keywords

video motion capture, text-based sound art, interactive installation.

1. INTRODUCTION

The creation of this installation came about from a desire to explore ideas of personal space in an abstract way. I wanted to de-humanize personal space so that participants could feel free to explore the various ways that we react to someone entering our personal space. To do this, I created an installation involving eight speakers, each with their own character. Each speaker was unique as possible in its shape, size and brand. They were arranged at different levels on different stands, or no stands at all, in a variety of locations in the space. Participants were invited to mingle amongst the speakers and the speakers spoke[1] with them. Some speakers were happy to be approached, some frightened, some impartial, some constantly changed their minds, some were simply odd, some existential and so on. Each speaker reacted in its own way to the movements of the audience.

---

[1] Not all the speakers spoke with understandable text but all were recorded phrase-like speech, nonsense or otherwise.

2. TEXT
2.1 Discussion of the text

The recorded text for this piece was the most important aspect. I wanted to avoid loops or simple reactions such as: if person A is one metre away from speaker M then say X, if person A is fifty centimetres away from the speaker N then say Z. In this regard, I needed a variety of similar text that could be randomly chosen based on various sensor data. Depending on the speaker, there were groupings of text that were triggered depending on whether there was motion in front of, behind, near, or very close to the speaker. There was an ambient grouping of text that was randomly played if there was no-one moving in proximity to the speaker.

To gather the text, I invited colleagues, friends and family to individually join me in a studio and asked them to speak as if they were: alone in a room; in a room and people were entering at a distance; near to people; very close to people; being approached from the front; being approached from behind.

I must take the space here to give a huge thank you to all of the volunteers who dove into this unusual recording session and fearlessly gave streams of wonderful text in a very abstract setting. I did have characters prepared for those who could not think of anything to say or got stuck, but for the most part the volunteers were amazing in their ability to jump into the situation. I was looking for a variety of characters and I think the piece was much richer for the eager participation of the people I recorded. Their eagerness offered much inspiration and energy to the piece.

That being said, I think the richness, diversity and oddness of the characters did create some confusion in participant's understanding of the installation. The text of each character was so engaging in and of itself, that many people did not realize that their movements altered what the speakers would say. This could have been due to how I organized the sound files, but I believe it was more to do with the inherit richness of the text as a whole.

This is really a question of how far I could suspend the audience's disbelief. I find that in interactive art, installation art and similar new media art forms the suspension of disbelief is much lower than in theatre or other more traditional art forms. Perhaps this is due to the fact that there is little frame of reference for most audience members, especially those unfamiliar with sound art, installation art and new media. Many people still do not know what sonic art is. This is a great challenge faced by artists using new media. How can we keep complexity and richness in our work without sacrificing accessibility? How can we define a context for understanding a

work that does not have a readymade context? These are outstanding questions for any artwork but there is a contextual haze around new media art[2](p1section2). This struggle for context by the audience is compounded by many new media art being focused on technical gadgetry as opposed to art.

For example, all of the characters in *Mixed Company* had the richness I mentioned above except for one. It was the one with the recording of my own voice. There were two reasons why the recording of my voice was unique. I was the only character who used exclusively one word, "hi" (I recorded myself saying "hi" in about 100 different ways). Also, I was also the only person who really knew how the final installation was going to take shape.

From observations and comments, the speaker with my voice was the one that most people actually wanted to talk back to. Perhaps the simplicity of my speaker's character was easier to relate to than the more complex and rich characters of the other speakers?

I could have had each character use only one word or phrase such as, "hi", "what", "how are you?", "come closer", "I love you", "go away", etc. but I believe I would I have sacrificed complexity of character.

I will take an example of a speaker with richer text. Some examples of what the speaker said were, "gather round everyone, gather round", "you're stupid", "you over there", "oh great, just ignore me", "sorry were you talking me?", "how's it going", "no, I wouldn't stand that close to me", "you can come a bit closer to me, I don't mind", "you're fine actually, that's quite nice what you're doing", "um", "ah", <whistling>, etc. This type of complexity offers a richer experience for the audience as they are continually surprised and they cannot easily discover how the piece works technically. It was my hope that they would consider and reflect upon their interaction with the artwork instead of the technical construct.

Perhaps a furthering of this installation could see it being split into two. One with simple text and one with complex text. Or, the speakers could be silent until people moved in proximity to them. This would make the technical aspects of the piece more transparent but again, would I be sacrificing complexity and richness?

I do not think an infinitely rich text or an overly simple one is the answer but somewhere in the middle. A more decisive and detailed categorization of the text combined with observation and listening to the comments of participants will help make the instillation stronger but educating an audience takes time and experience.

2.2 Influences and Inspirations on the text

The most direct influence on the text was Eric Belgum's *Bad Marriage Mantra*. This text is a series of phrases that can be spoken wholly or in parts. The text is meant to be spoken over and over again and the actors choose which whole or part of each phrase to say at any given time. The result is that a relatively small amount of text can generate a long dialogue by the random choosing of its parts. Belgum equates the text to music when he describes a fight he overheard that inspired the making of *the Bad Marriage Mantra*. The fight... "had a great deal in common with many musical and literary traditions: the use of intense but slightly varied repetitions coupled with sparsely chosen materials."[1].

Another influence was the theatre of the absurd. Authors such as Samuel Becket and Eugène Ionesco were excellent at creating nonsensically sensical dialogue. They were excellent at creating a dialogue about nothing but making it sound like something or vice versa. Silliness and allegory abounds in their work.

3. SOUND DESIGN
3.1 Speaker choice and placement
The speaker choice and placement was relatively straightforward. I chose eight unique speakers to further define the difference between the characters. I placed them on stands, the floor and a seat, to create different levels and variety.

As mentioned above, the speaker with my voice generally created the strongest urge for audience members to talk back to it. Another possible reason for this, which came from some comments, was that it was the tallest speaker and was the only one with its cones at eye level for most people (the next tallest being around 150 cm high).

One audience member suggested putting clothes on the speakers or somehow making them more human. Although this is an interesting idea, I feel it would take away from a core element of the installation. The point was that the speakers are not human, so we can feel free to do things like peer at them up close, hug them, touch them, stand right behind them, et cetera, exploring a spatial relationship that would be uncomfortable or inappropriate with real people.

One idea that I did not incorporate but I think would be an interesting extension of the installation is to have the characters move from speaker to speaker. For example, if a person came close to speaker A, the character in speaker A would run to speaker B or conversely; if a person approached a speaker A, a character from speaker B would run over to speaker A. This would make for a very different installation as each speaker would lose its individual physical identity and it would become more of a virtual space where virtual characters react with the real audience.

3.2 Editing, Grouping and Random Playing of Text (probabilistic algorithms and simulated intelligence)
The most important part in the creation of the installation was the editing, grouping and playing of the text. After receiving all of the wonderful texts from my volunteers, it was my job to cut them all up and group them into appropriate categories in a combination of intuitive and logical categorization. Generally, I tried to cut the text up into its smallest meaningful components. For example, I would not cut up a phrase into single words unless it was something like, "yeah" or "hmm". It was also an effective choice to cut some phrases short so that other bits of randomly chosen text had the chance to finish a sentence in a different way. Again, I tried to keep a balance between intuitive and logical choices.

I then grouped the text from each character into two or three categories. Some text was not particularly diverse (such as my "hi" text) and some ranged from short monolgues to various grunts and giggles. The former I would separate into two categories; far and near. The later I would categorize in three groups: either far, near and very close; or far, in front of and behind, depending on the content of the text. I used the word far to describe the group that included ambient or 'alone-in-a-room' text. This was the text that was played when the speaker did not sense any motion.

Once the bits of text were organized in this way, I could place them in the patch I made using Max/MSP to randomly choose the sound files and play them. The speed at which they were chosen was random as well. I tended to give more time between the choosing of the far text as opposed to the near text. This gave a sense of urgency to the near comments. However each character was unique so I adjusted the randomness to suit each speaker's character.

The combination of small bits of text and their random selection created a very believable simulation of intelligence. There was a life to the installation. If one spent five to ten minutes with the installation, one would hear repetition but because of the randomness, even the repetition was not fatiguing or boring. This connects to the comments by Eric Belgum about "slightly varied repetitions coupled with sparsely chosen materials" and how they can be very engaging.

Other possibilities to increase the randomness and variety of the text would have been to play the sound files at slightly different speeds or use speech processing software such as Hideki Kawahara's STRAIGHT's[6] speech morphing.

4. SOFTWARE
4.1 Motion Capture: sensors or video?
The primary criteria in choosing the motion capture technology for this installation was that the audience would not have to put on any special garments, headsets, gloves, sensors, shoes or the like. I wanted the audience to walk into the installation space and 'magically' trigger the sensors.

The motion capture technology I ended up using came about from financial restrictions but revealed interesting capabilities that I may not have discovered had I gone the more expensive route.

At first, I wanted to use proximity sensors such as infrared or ultrasonic, but I needed a minimum of two sensors per speaker (preferably four) to get data from all sides. That made a total of 16 to 32 sensors not including a multiplexor and other peripheral electronics. As my budget was little more than zero, I sought another solution. I already have a computer and I had access to all the wonderful SARC facilities which included a digital video camera. With a camera, Max/MSP/Jitter and some crafty programming, I came up with a solution that was both better and worse than the sensor option. I describe the pros and cons below.[2]

---

[2] The pros and cons I describe are directly related to my project and my personal experience. For a more general discussion of motion capture please peruse the many articles on the subject from NIME, ACM, ICMC and more. I refer the reader to

The pros of the ultrasonic and infrared proximity sensors are that varying degrees of light do not alter their functionality unless it is infrared light in the case of the infrared sensors. Another pro of these sensors is that you receive continuous data from the maximum to minimum value whether the person is in motion or not.

The cons of the proximity sensors are that they are single axis and have a fixed range. This can be a severe limitation if you want to sense proximity all around an object or detect motion in unusually shaped areas. The electronics involved can be 'buggy', unreliable and less durable especially for long-running installations.

The pros of video motion capture are that you can choose to have the sensing technology be invisible. It is more adaptable to varying kinds of motion detection. It is durable and reliable. A huge pro for video motion detection is that you can easily choose the area where motion is detected.

The cons of video motion capture are that it is affected by varying degrees of light. Shadows are a problem so the direction of the light source is constrained as well. This can be a severe limitation in a small or awkward space that is difficult to light. Luckily this was not the case for this installation. It is more difficult to sense proximity if the person is not moving.

I did not explore the capabilities of an infrared camera which could be a better solution than a standard camera setup but even that is affected by certain types of lighting more so than infrared sensors. Ultrasonic sensors are not effect by light at all.

4.2 *Mixed Company's* motion capture software
The most useful aspect of the motion capture patch made for *Mixed Company* was its adaptability. I was able to easily focus in on the various areas where I wanted to detect motion. I could choose whether I wanted to detect motion in front of, behind, near or very close to an object with ease. I could change the area of focus as well. Essentially, I created digital zoom and the zoom could take any shape. I could cut out or zoom into any area on the screen and the area could be the shape of a half-circle or a triangle, or a donut or what-have-you and only detect the motion in that area. I found this to be an extremely versatile approach to motion detection. The software I used to create the patch was Max/MSP/Jitter.

For the motion detection, I used frame subtraction where the current frame is subtracted from the previous frame and the difference is displayed. Any change in any pixel is displayed and this is the motion. The limitation of this is that if someone stands still they are invisible to the system. In the context of the installation it worked well because when a person was still the speakers eventually went back to there 'alone-in-a-room' text which often consisted of phrases such as, "Helooo?", "Is anybody there?", "What am I doing here?", etc. Then, when the person moved again to leave the proximity of the speaker, the text would often beckon them back with phrases like, "What are you doing there?", "You're stupid" and/or with an increased sense of urgency in what they were saying.

---

cv.jit, Jitter and Eyesweb for examples and discussion of video motion capture techniques.

Another motion capture technique I would like to try in the future is background subtraction as opposed to frame subtraction. Although there are other issues with background subtraction, it might solve the con of not being able to sense a person in the space if they are not moving.

5. CONCLUSION

Overall the installation was quite successful. The interactivity could have been clearer but the issues described above hint that perhaps this is more complicated then simplifying or better defining the interactivity but a combination of this and the educating of audiences. The technology was very robust and adaptable with some limitations such as light change and an inability to sense stillness. The inability to sense stillness did not alter the effectiveness of the installation but the restrictions on light did hinder a dramatic lighting of the installation. However, with some more time or a lighting designer a lighting rig that was both aesthetic and functional could have been setup. The system remains limited in this regard.

6. ACKNOWLEDGMENTS

A huge thank you to all the voices: Chris Chong, Javier Jaimovich, Maria Ponce, Donal O'Brien, Mac Dodge, Julia Dodge, Heather Richards, Bertram Knappitsch and Wolfi. Thank you to SARC for its staff and facilities. I would particularly like to thank my lecturers, Paul Stapleton, Michael Gurevich, Ben Knapp. And Chris Corrigan for his technical assistance.

7. REFERENCES

[1] Belgum, Eric. http://www.freewords.org/biennial/artist/badmarriage.html. Viewed May 25th, 2008.

[2] Birchfield, Phillips, Kidané and Lorig. Interactive Public Sound Art: a case study. Proceedings of the 2006 International Conference on New Interfaces for Musical Expression, Paris, France.

8. LINKS

[3] EyesWeb http://www.infomus.org/EywMain.html

[4] cv.jit http://www.iamas.ac.jp/~jovan02/cv/

[5] Jitter http://www.cycling74.com/products/jitteroverview

[6] STRAIGHT, a speech analysis, modification and synthesis system http://www.wakayama-u.ac.jp/~kawahara/STRAIGHTadv/index_e.html

9. SIMILAR WORKS

[7] Legrady, George. Sensing Speaking Space, http://www.mat.ucsb.edu/%7Eg.legrady/glWeb/Projects/spsp/Spsp.html

[8] Snibbe, Scott. Shy, 2003, http://www.snibbe.com/scott/screen/shy/index.html

[9] Adam, Frank. Shadow, 2003, http://www.mine-control.com/shadow.html