Speechless - a network experiment with rhythmic text

Alex McLean

May 5, 2007

1 Introduction

This project was borne primarily from two interests. Firstly, a strong research interest in relationships between speech, text and instrumental sounds in the context of music. Secondly, in the more technically motivated possibilities of interacting with audio streams via a webpage, found through experiments with the HaXe language.

In this report, relevant research background will be highlighted, followed by the original design specification that it inspired and technical details of its refinement and implementation. The results of this project (so far) will then be summarised followed by conclusions made.

2 Background

Human perception of sound is strongly influenced by the physical articulation of speech, as shown by the McGurk effect [5], where the sight of a mouth moving drastically modifies our perception of an accompanying sound.

This speech-sound connection may also be seen in our use of onomatopoeia, where we are able to hear a sound (such as an explosion), reproduce some estimation of it with our vocal apparatus and transcribe it with a vocable word (such as "boooomshh"). There are many onomatopoeic words to be found in an English dictionary, such as "boom", "tinkle" and "quack". These words and the sounds they describe are far from universal, for example different cultures attribute quite different onomatopoeic sounds to the same animal.[2]

Kohler[4] describes an experiment suggesting a strong connection between words and visual form. When presented with an angular shape and a curvy shape such as in Figure 1, and given the meaningless words *Maluma* and *Takete*, almost all name the angular shape Takete and the curvy one Maluma. This relationship between parts of words and meaning is known as *phonaesthesia.*

All of this indicates strong interrelations between sounds, the structure of words, and abstract form. It is no surprise then to find these relationships extend into the domain of music. Indeed, there are rich parallel histories of text-based notation systems in many cultures. The term *vocable* is used to describe syllables which notate instrumental actions within such a system. In Hindustani *Tabla*, rhythms described with vocables are known as *bols*, where a particular vocable represents a drum stroke within a particular tradition[3]. A loose relationship can be seen between the relative structures of bol syllables and the technique of playing them. This relationship is even

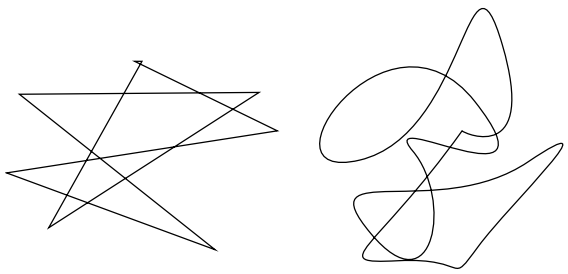

Figure 1: Maluma and Takete - which is which?

```
Fümms bö wö tää zää Uu, pögiff, kwii Ee.
Ooooooooooooooooooooooo,
dll rrrrr beeeee bö
dll rrrrr beeeee bö fümms bö
    rrrrr beeeee bö fümms bö wö,
          beeeee bö fümms bö wö tää
                 bö fümms bö wö tää zää
                    fümms bö wö tää zää Uu
```

Figure 2: Excerpt from the introduction of *Ursonate*, by *Kurt Schwitters*

stronger in *Canntaireachd*, an ancient Gaelic system of vocables[1], commonly associated with the bagpipes[1]. The vocables notate not only pitch but also timing and ornamentation. With both bols and canntaireachd, vocables are used for communication in pedagogy and during performance as refrain, as well as mnemonic aid and notation.

We should also consider the work of sound poets such as Kurt Schwitters, whose sonata *Ursonate* (Fig. 2), a poem of rhythms defined by words without extrinsic meaning, is a stunning reminder of the expressive power of speech. Indeed, this piece is perhaps the greatest inspiration for this project.

3 Design specification

The idea was originally specified as follows;

- Person visits a webpage and immediately sees and hears others
- They hear synthesised words spoken, and sees text of those words, repeating phrases
- The words may or may not make sense but will be spoken in a machine-like English voice
- They see a text box and are invited to contribute some words
- They type "this is strange. what is this?"
- Immediately they see this added to the other phrases. The . and ? are turned into pauses.
- A few seconds later they hear the words mixed in with those of other players.
- They explore further, adding more periods and playing with different word structures, making up new words with interesting outputs
- They are also invited to buy a DVDs worth of compressed audio from the duration (two weeks?) of this stage of the project which will be dispatched on completion for £15. It will somehow come with lyrics metadata containing the strings used.
- Only 5 people can play at any one time, with a queuing system. *The final system had a maximum of 3 people.*
- However up to 30 people may listen to the stream, and 70 more to an ogg stream.

3.1 Differentiating voices

In order to make it possible for players to quickly identify their own 'voice' within the system, the number of concurrent players was set to three. Their input boxes were shown across the page from left to right, and the sounds they made panned accordingly. For example the player on the left would hear their sounds only on the left channel. Further, the leftmost sounds were pitched lower, and the rightmost higher. I did not go further in adding differentiating effects in order to keep focus on playing with the structure of the words in order to produce varied sounds.

[1] Vocable words are those that describe sounds which may be spoken or played. They generally have no meaning beyond representing a sound.

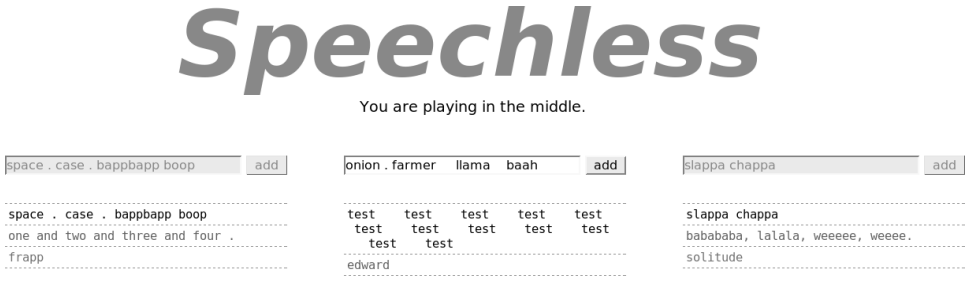

Figure 3: Main section of the interface

4 Implementation

4.1 Interface

To allow myself greater chances of being surprised by the players, I decided to give them the bare minimum of instructions, although they were fleshed out a little in response to frustrations aired during the beta phase. The main section of the interface was as shown in Figure 3.

If all the slots were taken, all the text boxes would be greyed out and disabled, and the player's queue position was shown under the Speechless logo. If they were next up, they were told to get ready, and finally when they had control of a text box, it was enabled and the text changed to (for example) "You are playing in the middle".

The instructions read as follows:

1. Make sure you have flash 8 or higher, your sound is on and turned up
2. If the boxes are all in use, wait in the queue until one becomes available
3. Type in the box. Make up words that sound nice. Add gaps between words with spaces and punctuation.
4. If others are waiting, you will get one minute to play before rejoining the queue
5. If you're alone, and the other voices are getting annoying, get a friend elsewhere on the Internet to join in.

Below these instructions was an invitation to buy a DVD containing compressed audio from the two week run of the project, followed by a simple feedback form.

4.2 Network

Many bits of software were connected together to realise this project, as shown in Fig. 4. Both javascript and flash (actionscript) ran in the browser, and a Perl script, the Festival speech synthesiser, the datadirt sampler and icecast media streamer running on the server. Both the javascript and flash were compiled from HaXe. An outline of each part of the system is shown below.

4.3 HaXe

HaXe is free software which allows a programmer to use one language that compiles into several 'target' languages. I used HaXe to produce both javascript and Flash code that ran in players' browsers.

http://haxe.org/

4.3.1 Flash

A flash object was needed primarily for playing the mp3 stream from the icecast server, so the browser could begin playing the stream as soon as the webpage loads, without requiring the user to launch an external media player.

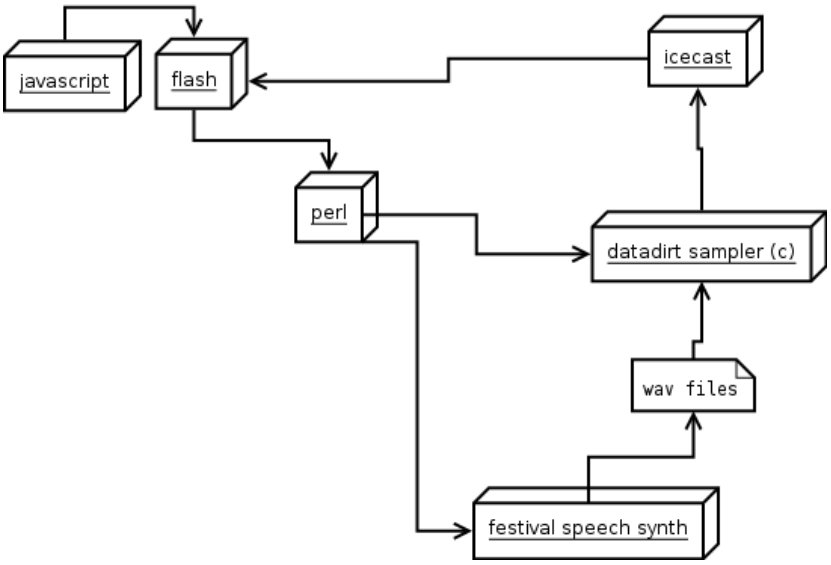

Figure 4: Network diagram

The flash was also used to maintain a client-server network connection, passing messages back and forth between the Perl server and the javascript.

Javascript The flash object described above was invisible, all the player saw was a text based interface rendered with HTML. Javascript was used for all the user interface features, showing queue or player position, taking input from the player, showing the input of other players, and so on.

Perl server This is a Perl script responsible for chopping player input into words and sending them to the speech synthesiser, and later triggering playback of the synthesised words in the sampler. It is structured around a heartbeat, so that it may trigger the sounds at the right time. It handles all communication with the HaXe software in the browsers, keeping them informed of what is playing where. It also maintains a timestamped log of all player input.

Festival speech synthesiser Festival is free software which among other things allows text-to-speech synthesis. For this project I used the *text2wave* commandline tool, which offers a simple interface for producing audio files from strings of text.

http://www.cstr.ed.ac.uk/projects/festival/

Datadirt Datadirt is free software, written by myself, for triggering of efficient live playback of samples with effects. For this project I modified it to encode audio output in compressed MP3 format and send it directly to the icecast media streamer.

http://yaxu.org/datadirt/

Icecast Icecast is free software for multicasting audio and video streams.

http://icecast.org/

5 Results

5.1 Beta

I initially ran a 'beta' version of the project, inviting colleagues and friends to use and give feedback.

By far the biggest criticism during this period was that it was irritating to use if the other two 'slots' were not in use. This was because utterances of previous player remained, looping indefinitely until the next player comes along. I countered this by making a small change, where the volume of an utterance reduced exponentially over time, finishing off at one quarter of its volume. A compressor was then applied to bring the overall volume up again. Then a player would always find some utterances when they came to speechless, but their own contributions would mostly drown them out once they started playing.

I also discovered and fixed a bug in the datadirt sampler, where it was failing to close files and therefore running out of filehandles after a few hours of use. Once fixed the whole system ran continuously without crashes.

5.2 Launch

The project was launched on Resonance FM, a radio art station broadcasting from London, UK. Several live performers were in the studio, jamming along with the speechless project. Listeners were invited to contribute to the music not only by typing into the speechless website but also phoning a number and making noises down the line. Players were active in Speechless throughout, and the experiment as a whole was deemed a success by the organisers.

For the rest of the project, speechless ran on its own. I mailed several related mailing lists, inviting people to participate, and presented it at the *dorkbotlondon* electronic art forum.

5.3 Use

I logged every utterance, with a timestamp, for later analysis and reconstruction. There were over 12000 utterances in total, although many of them were repeated. To give a flavour, here are some excerpts from the log with commentary.

5.3.1 Creative use of space

```
f           sf           sf             ff            sf
f           f           sf             ff            f
f           f           f             ff            f
```

Space didn't do anything in the beta, but many people used it to try to space out sounds. In response, the final version treated each pair of spaces as a rest between sounds.

5.3.2 Cut and paste

```
Dreamy smooth kenji, kanemasu panamusica fzymc publisher. Yields andor companies ultimate attractive prices spreads
```

Many players cut-and-pasted bits of text in. Bits of text from the webpage, pieces of poetry, the Irish national anthem and (as above) excerpts of spam. I think the requirement of javascript and flash would have prevented any real spam bots from contributing, which is perhaps a shame.

The maximum that was pasted in was a 2,500 word dictionary. Impressively, the system didn't crash, although that would have been annoying for the other players present.

5.3.3 Soap box

Contributions to speechless were anonymous, and so the players are only identified here by their position 0, 1 or 2:

```
1: Therefore I must tell you, first of all, what Anarchism is not.
0: brr
1: It is not bombs, disorder, or chaos.
1: It is not robbery and murder.
0: You are playing on the left hand side.
2: mmUUURMMmmm
1: It is not a war of each against all.
1: It is not a return to barbarism or to the wild state of man.
1: Anarchism is the very opposite of all that.
2: uhhhck kuck fal qiiiiiieeeek
```

On a couple of occasions, someone started pasting in lines from political treatise. As in this example, the other players did not seem to be taking the message on board.

5.3.4 Conversations

Occasional attempts at conversation were frustrated by the players being shuffled around the slots by the Speechless server, and apparently getting distracted by the sounds their words were making. In this interchange Marcus was particularly frustrated, failing to get an instant messaging ID from Anne.

```
0: whre do you from???
1: paris
0: bien
0: et votre nom???
0: je suis Marcus
1: bonjour marcus
0: je suis brasillien
0: tell me your name
1: a n n e
```

5.3.5 Percussive sounds

```
boom, tsik boom, boom tsik tsek
boom chaka laka boom chaka laka boom chaka laka boom
AK AK Oh OH AK AK OH OH Aka laka boom
ka ka ma mo me oh my
bap bap bapbap . bays . bays? . couches couches couches couches
booom tchi boom tchi tcha tuum tcha
tinsh tish top tap! takkatakka . tappatappa? .
```

Perhaps the most successful attempts at music with Speechless involved short, percussive sounds.

5.4 Feedback

A large number of people left short messages via the website, which I include in Appendix A. I also received some more detailed messages by email, some of which are included with comments below.

> "I'm assuming this is looping three synthesized samples of the last pieces of text typed into the fields, but it sounds as if there are more than three voices going on. That might just be a shortcoming of my speakers though. It's a shame, but I can't hear more polyrhythm going on which is what I'd expect if this is the idea."
>
> *Adrian Ward, eu-gene generative art mailing list*

I think the problem Adrian is reporting is partly the fault of words overlapping each other during the beta test, and partly due to the way the voices are pitched. Overlapping words was reduced in the final version simply by running the system at a slower rate.

The problem with pitching was due to my making words playback with a higher or lower pitch simply by playing them faster or slower. Of course this makes them shorter or longer as well, making the onset of a sound perceptually earlier or later. This difference of onsets between voices could be another reason why polyrhythms were not more readily apparent. One solution could be to apply an volume envelope to the resulting sound with a fixed length onset, but I did not have further time to experiment before the end of the beta phase.

> "I was playing alone for a bit but found that you can open a couple of browser tabs to open two (I didn't try three) sessions at once, and play on your own :) It was much more fun when someone else did come along though."
>
> *Julie Freeman, private email*

Somehow I hadn't actually thought of opening the web interface multiple times until Julie emailed this. It works very well.

> "I'm too old for that web page to be interesting."
>
> *Randal Schwartz, (void) mailing list*

Nice for Randal to let us know.

> "i didn't do a thorough test, but what i did seemed fine. but is this a game? i'm asking because it says "you play on the left" and the right box was greyed out."
>
> (later)
>
> "ah, now that there seem to be more people connected i get it!"
>
> *Pedro Figueiredo, private email*

This makes clear how important interacting with other players is in understanding the project.

6 Conclusion

The feedback was pleasingly mixed, split between disgust, confusion and love. This range of response is gratifying on an artistic level, however the lack of universal appeal however meant the project did not fare well on web 2.0 social bookmarking websites such as reddit.com. Nonetheless, a good many people played with the system, as can be seen from the amount of feedback in AppendixA. It would be interesting to have been able to match response to the project with the cultural origin of the player; there did seem to be an upturn in use of the word *penis* while the project was discussed on the *livejournal* website.

The project was not a financial success - only one person wanted to buy a DVD, and was instead happy with a free copy of the sourcecode with the timestamped log, so they could generate their own.

As the designer/programmer, I gained a great deal from this project. On a practical level I gained an fairly novel, technically successful way of allowing multiple people to play with my text-to-music interfaces. I will undoubtedly make more use of in the future, particularly in the text-to-synthesis system I am developing as part of my MSc thesis.[6]

A Full feedback log

The feedback was received via a form in the website in the order shown, more recent feedback coming later.

- This is very interesting alex. I'm not quite sure where you're going with it, but I like it very much. I'm dancing around my kitchen.
- This is truly a pointless project.
- Perhaps a change of meter sometimes. It gets very repetitive. otherwise, it's great! julian / www.hypo.io
- it would be nice to have the ability to adjust the speed of the samples
- Any chance of a quick note to say what this is doing? - Greg
- Great stuff!! Really successful I think
- reminds me of brian eno's revox loops (which i sorely imitated with two ordinary tape recorders and regular cassettes). excellent.
- nah - too slow.. interest fading quickly
- Crazy and anoying!
- help me
- mystifying , what is happening ?
- This is really, really weird.
- nice

- This is fantastic.
- No sound produced on IE6 running on XP-SP2
- it works real well! totally fun for the user, collaborating on something with a stranger . definitely develops its own dynamics! mariax
- it works real well! totally fun for the user, collaborating on something with a stranger . definitely develops its own dynamics! mariax a longer history would be nice, so you can see all you wrote, or just more lines, the image of the text is interesting, a sound poem
- this *needs* to be streamed so i can listen to it in itunes :)
- Irregular pauses between loops please? Interesting project!
- great fun interaction. played for about 30mins until accomplices ran out.
- maybe you should let users harvest great configurations for you, with a "save this set" button. or "save last 5 mins". The latter could just save a timestamp and you could automate playback when no one was using it (an attract mode).
- it's terrible being locked in to one slot when the others have been left with uninteresting or very noisy settings. oh and this feedback box gives unclear feedback if you use it more than once.
- Its like a bad acid trip!
- I tried this at work and now at home and it doesn't work. IE 7 WinXP Flash 8 John Clavin
- awesome. is it possible to change the pitch/speed somehow?
- An interesting symbiotic mixing of computer synthesized voices talking nonsence
- HOw does this work anyway?
- it sounds a bit messy to me ....
- kewl
- I struggle to 'hear' the individual words. Varying the word length doesn't appear to influence the length of the looped sound proportionately. Perhaps the loops are too short? I kind of like it, but to be honest the sounds produced are far from pleasant, and soon get very annoying (is this intentional?)
- it's addictive
- fun! should have the option of playing all three fields if no one else is online though
- i want now
- 1. What's the goal of the game? What are we playing for? I've absolutely no idea what to type in the boxes above. 2. I'm unable to hear any sound even though my output volume is set to maximum (my machine: Windows XP, Internet Explorer 7, Flash 9)
- i think i got there when there was nobdy else to talk to, but was quite interesting and the music a little bit scary. not sure about having to delete the text each time. trying to see what gives it rhythm, the types of words letter or the rythm typed, but felt a bit silly so not truely explored. like alot. need a playmate.
- Very entertaining and engaging. Quite frustrating not to be able to change the centre and right as well as the left - I guess no-one else was using it when I was. Would be nice to be able to remove things from the list. Geraint
- this is great! i'd like a version of this that could run on a local area network! .
- suggestions: - accept words as url parameters so a combination could be saved as a link - add each combination as a history link so visitors could review other combinations - allow thumbs up/down voting on combinations - choose random word button - vary pitch/tempo? - volume controls w/ mute - small visualization - short description/instructions It wasn't clear that I could add words using the one open field. Interesting and fun start, thanks!
- Instructions showed up on second load of the page. Pitch diff showed up on 2nd page load. I see now.
- Seems to start with the right word then lose a syllable or two... with long words
- needs control over the voices.
- 1st thought: weirddd

- vary the rhythmic patterns
- please put in a volume control feature!
- i loved playing it's a wonderfull piece
- seen it at dorkbot, it's really great maya
- It's a little odd.
- wooooooooooooooooooooooooooooooooooooooooooooooooo!!!!! .
- This is fun! thanks for sharing this idea. love, Eleanor
- great! great! great!

References

[1] Chambers, C. K. *Non-lexical vocables in Scottish traditional music.* PhD thesis, 1980.

[2] de Rijke, V., Ward, A., Stockhausen, K., Drever, J. L., and Abdullah, H. Quack project cd cover notes, http://quack-project.com/, 2003.

[3] Devdan, S. Tabla - technique. In *Grove music online (accessed March 2007)*, L. Macy, Ed. Oxford University Press, 2007.

[4] Kohler, W. *Gestalt Psychology.* Liveright, 1938.

[5] Mcgurck, H., and Macdonald, J. W. Hearing lips and seeing voices. *Nature 264*, 246-248 (1976).

[6] McLean, A. Livecoding: languages and models - msc thesis. 2007.