Generating video in realtime with recurrent neural networks

Douglas Bagnall <douglas.bagnall@catalyst.net.nz>

Disclaimers

✗ ~~photo-realistic~~

✗ ~~useful~~

✗ ~~state of the art~~

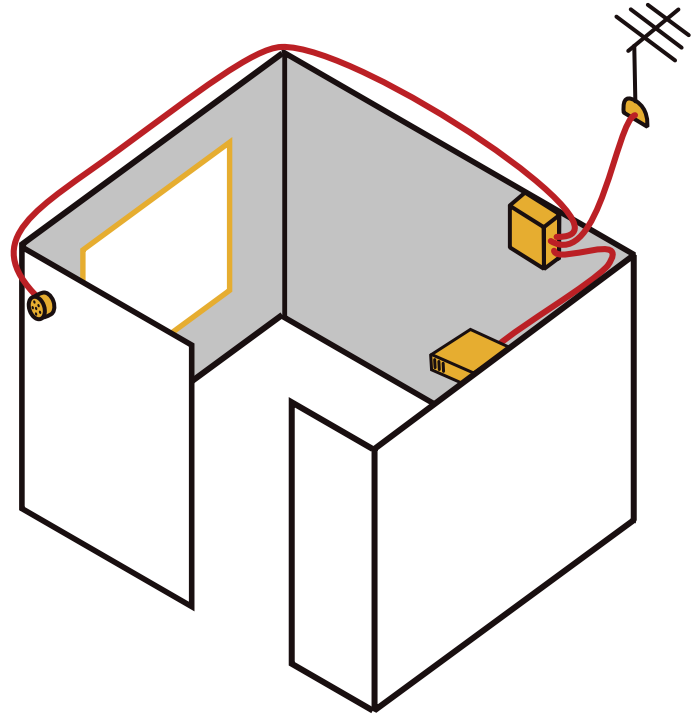
✓ made by mistake

✓ “works”

A proposal (2011/2012)

The computer will:

- listen to conversations of approaching people
- create related video
- present the video
- *not* tell them

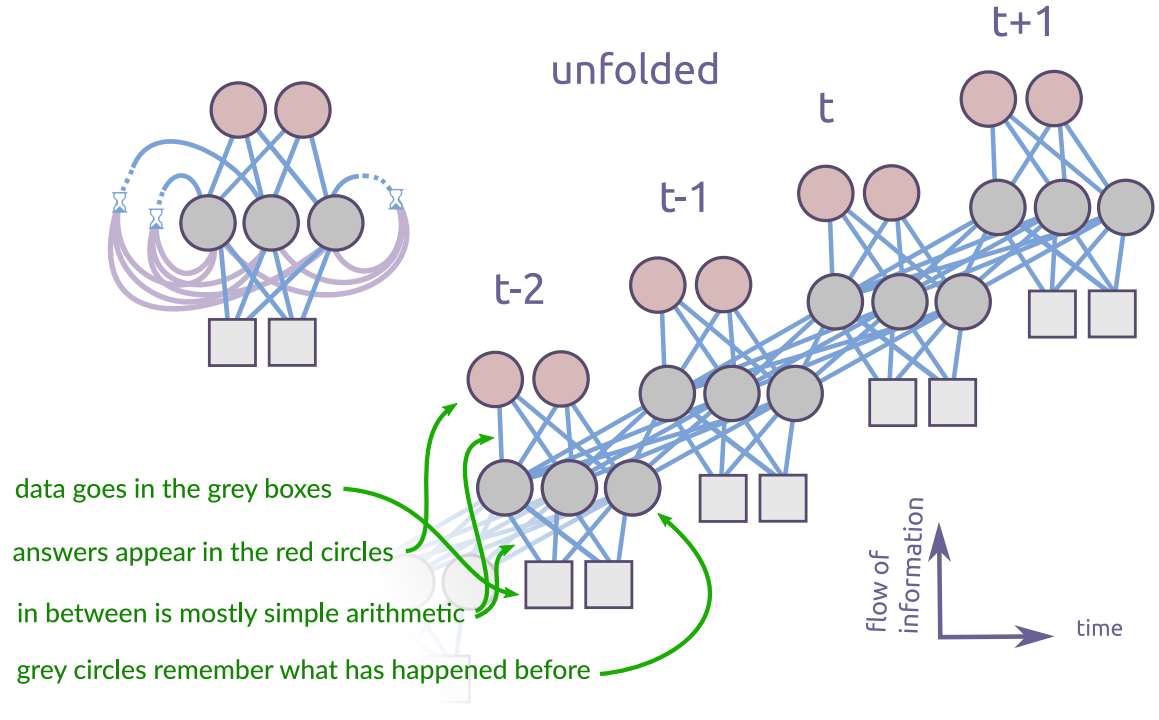


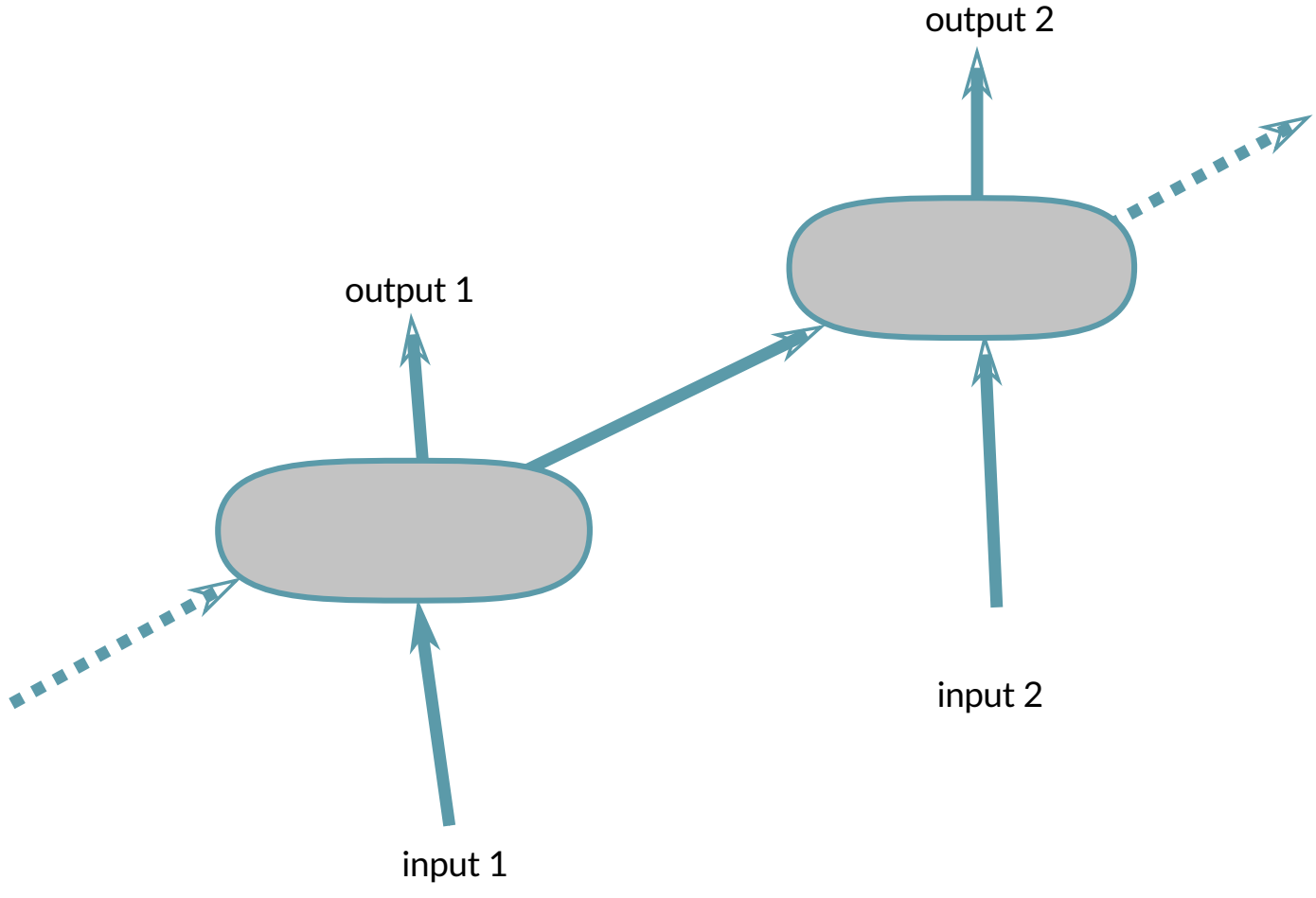
Proposed technology

- Speech recognition: PocketSphinx *not for NZ English*
- Video collection, tagging: broadcast TV, subtitles *no*
- Video creation: **???**

speech recognition investigation
lead me to
recurrent neural networks,
newly re-used in2011 for language modelling

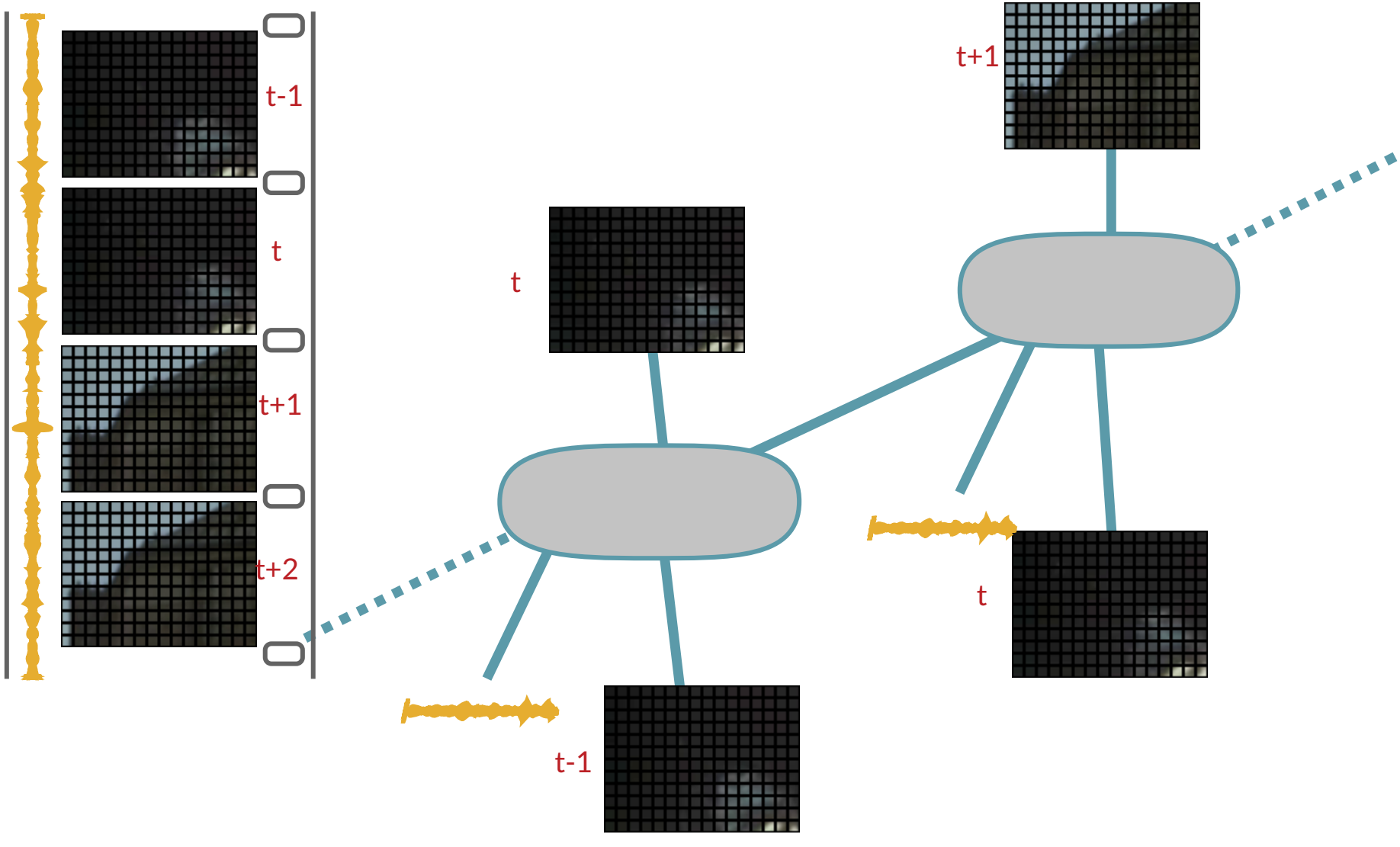
A simple (Elman) RNN

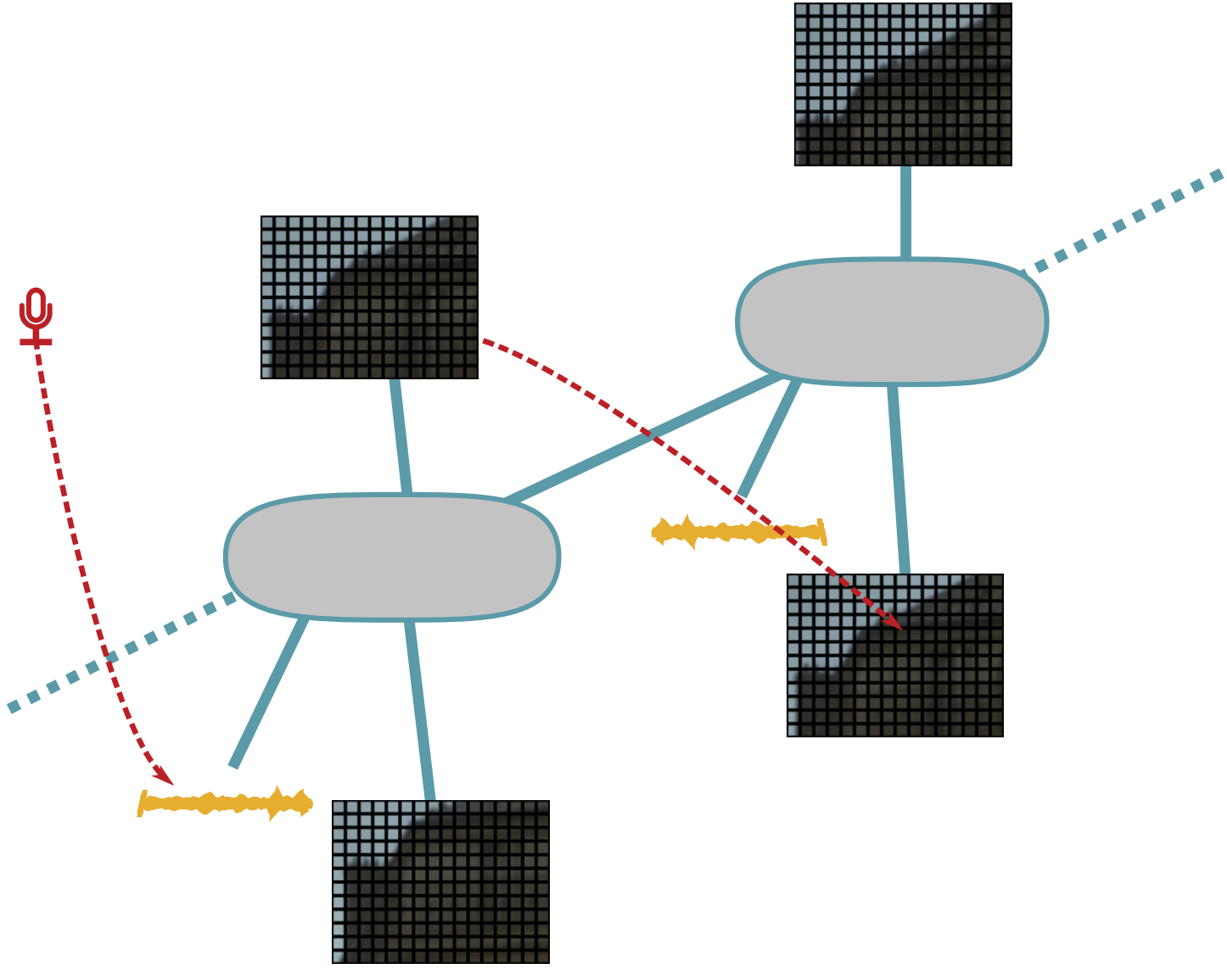



output 2
output 1
input 2
input 1

The deadline arriveth

- no speech recognition
- no labelled video
- no video creation algorithm
- an interest in *RNNs*


t-1
t
t+1
t+2
t+1
t
t
t-1

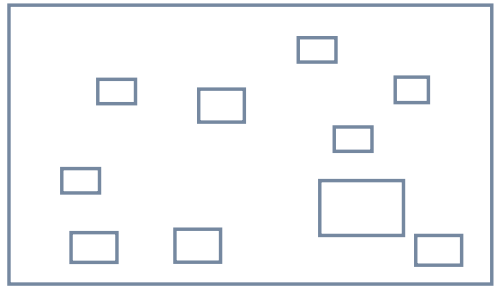


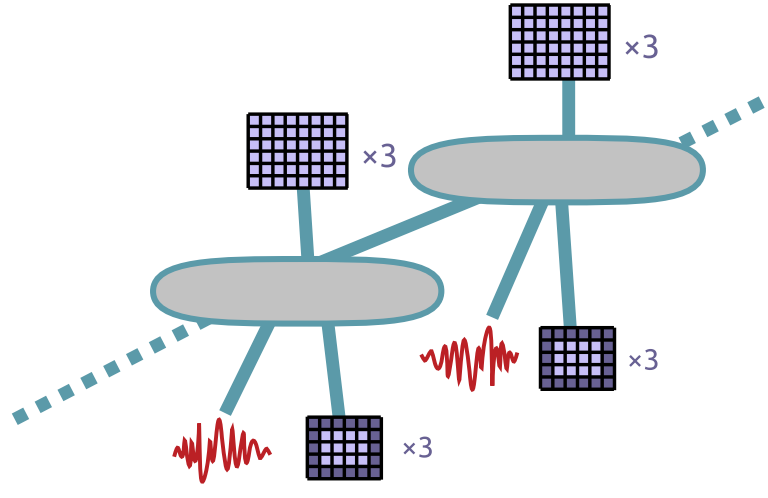
RNN generation of full video

- expensive (thousands or millions of nodes)
- needs millions of frames of training data

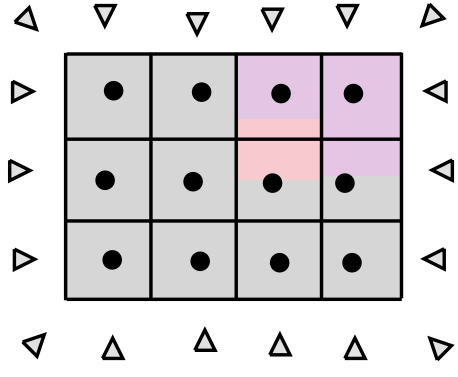
Recur (2013)

- Assume video is self-similar across scales
- Learn video in association with audio
- Output increases resolution
- recursively applied

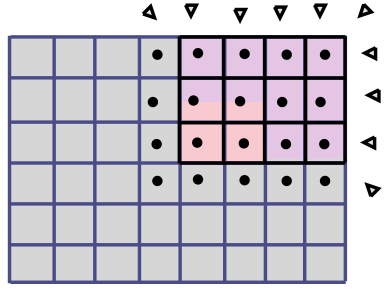



×3
×3
×3
×3

predict 8×6 from previous 8x6 resampled as 4×3.

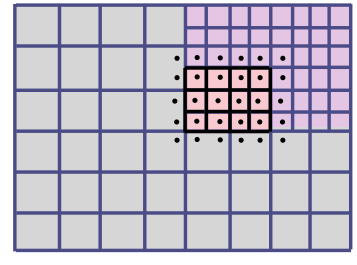


predict each quarter of 16×12 from previous 8×6.

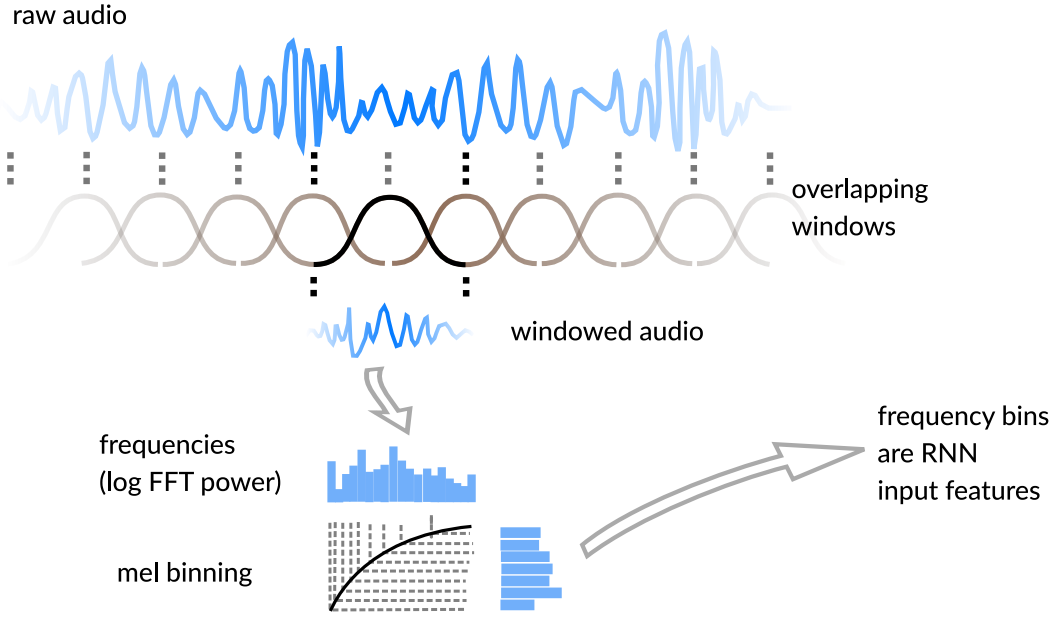


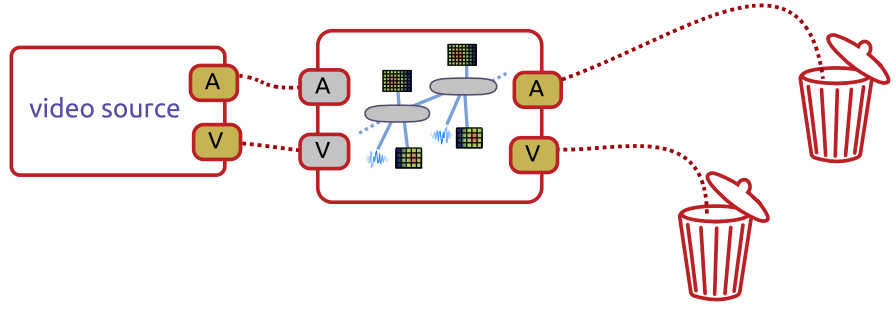
predict 32×24 segments from previous 16×12.

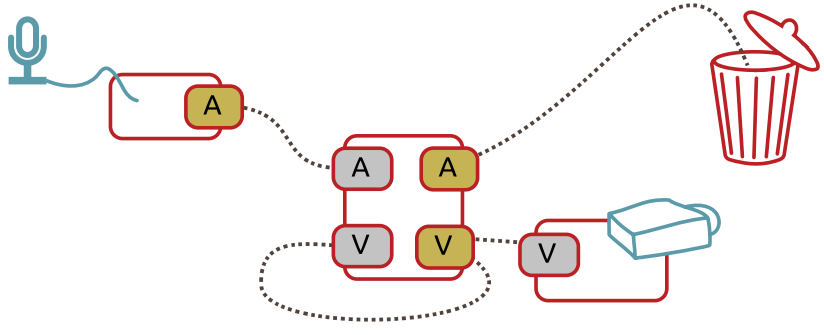
etc.

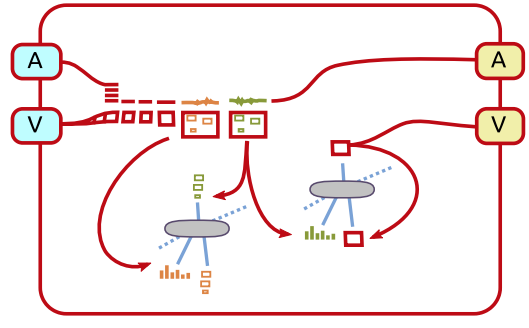


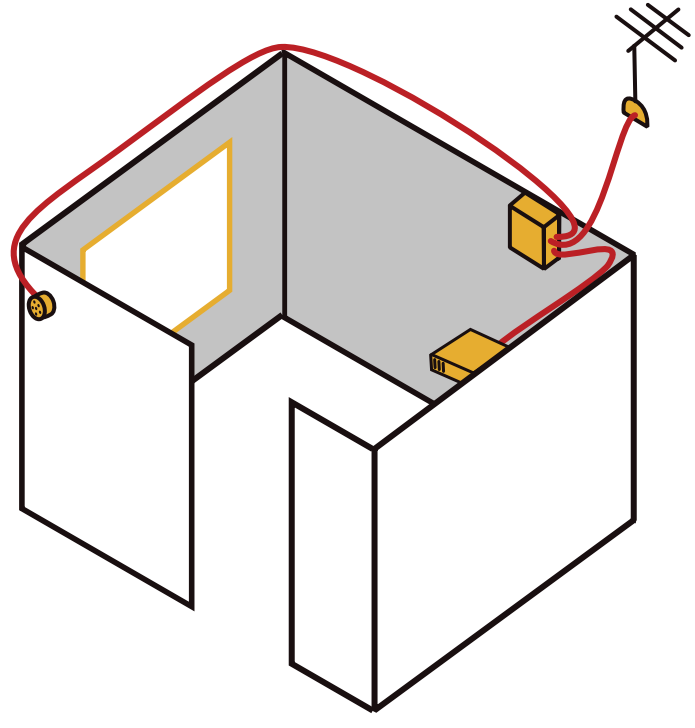
(I don't know why).

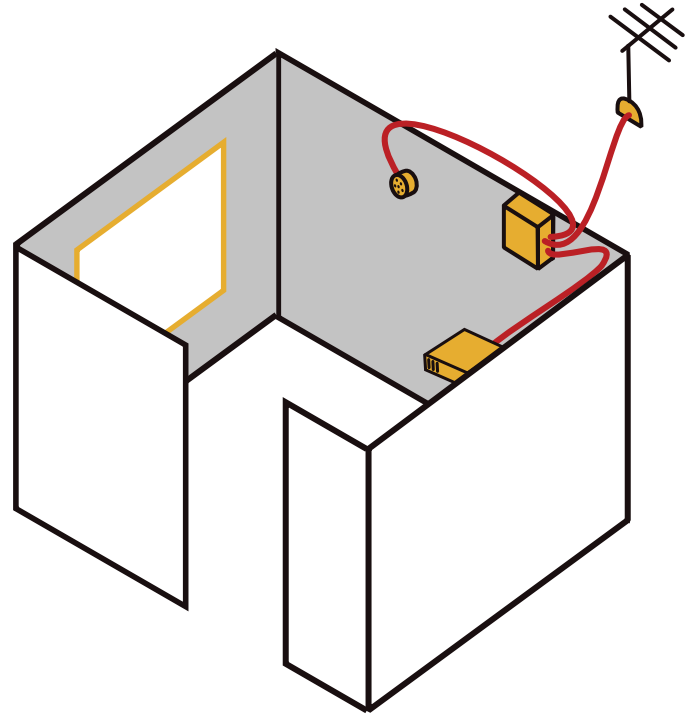

raw audio
overlapping
windows
windowed audio
frequencies
(log FFT power)
mel binning
frequency bins
are RNN
input features

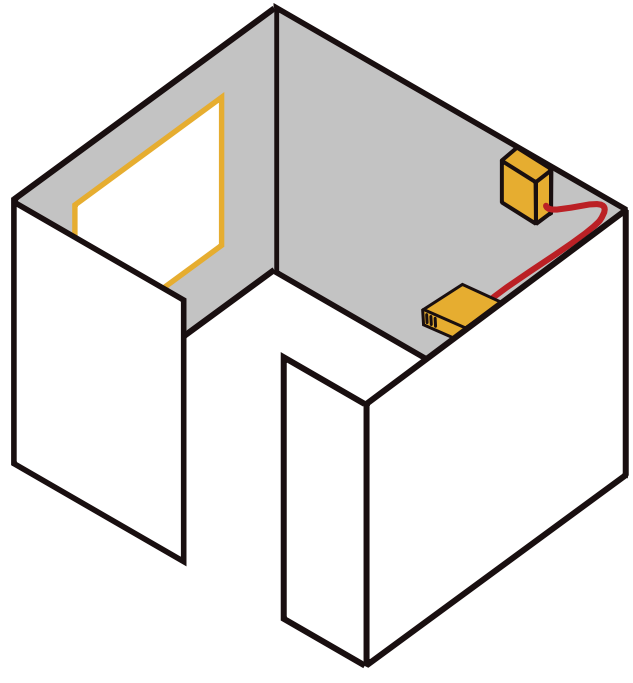

video source
A
V
A
V
A
V


A
A
A
V
V
V


A
V
A
V







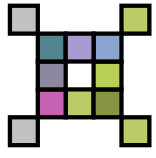
this slide is to remind me not to forget the demo

RNNCA

Cellular automata with:

- learnt rules
- internal state

input patterns can vary



and the other demo

What I would do now

- different generators at different levels
- one-bit networks
- muck around until the last minute, again

The code

- C and some Python
- Gstreamer framework
- LGPL/GPL 2+

Faster matrix operations than ATLAS

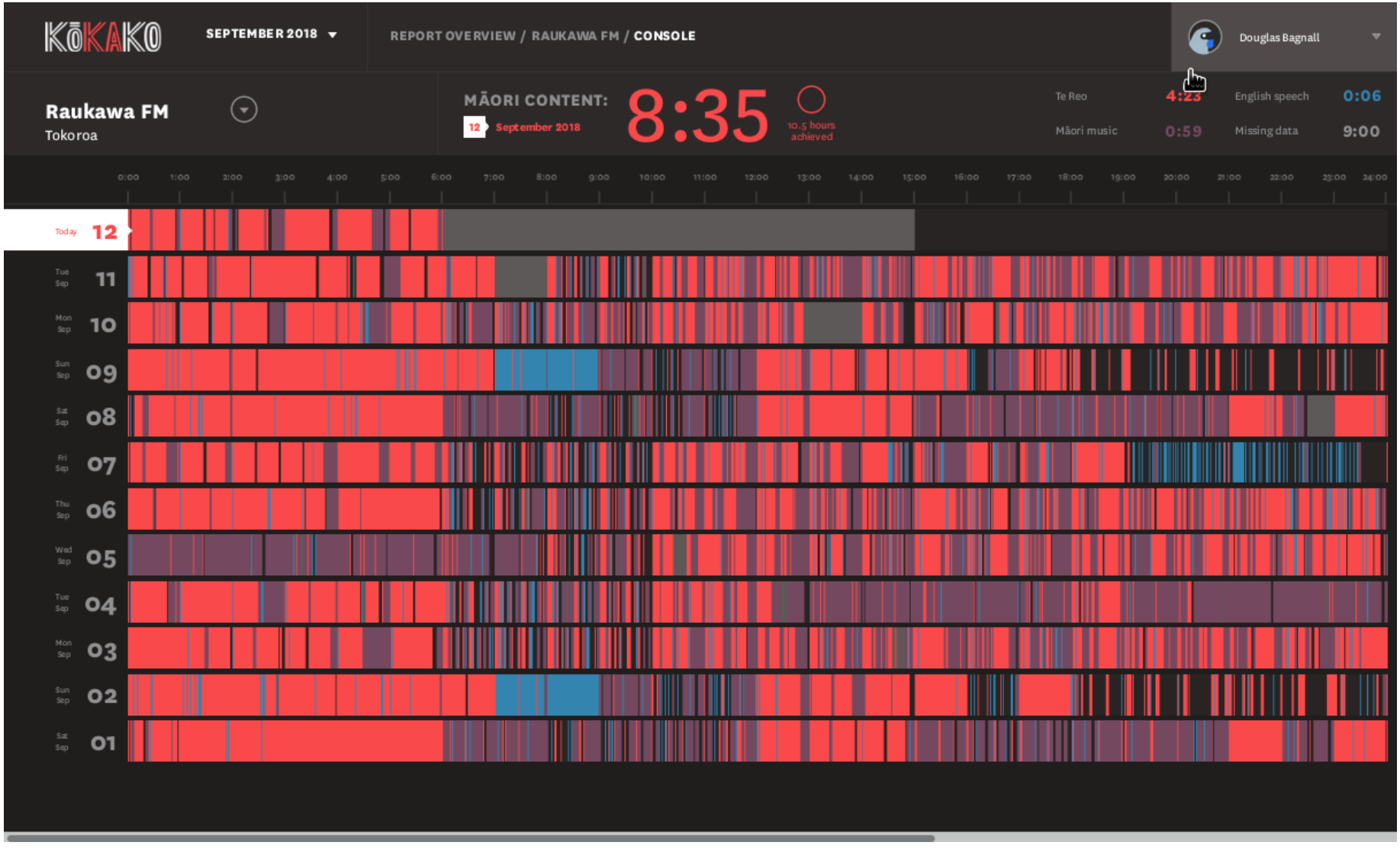
- because the data is known
  - avoid multiplies by 0, 1
- all matrices are aligned for SIMD
- allocations at start-up

```
#define ASSUME_ALIGNED(x) \
      (x) = __builtin_assume_aligned ((x), 16)
#define ASSUME_ALIGNED_LENGTH(x) (x) = ((x) & ~3ULL)

static inline void
foo(const float *restrict inputs,
    size_t n_inputs,
    const float *restrict outputs,
    size_t n_outputs)
{
   ASSUME_ALIGNED(inputs);
   ASSUME_ALIGNED(outputs);
   ASSUME_ALIGNED_LENGTH(n_inputs);
   ASSUME_ALIGNED_LENGTH(n_outputs);
    /* now GCC knows it can be fast */
}
```

Recur RNN reuse

- identifying the languages spoken on the radio
- identifying anonymous authors
- identifying bird calls


KōKAKO
SEPTEMBER 2018
REPORT OVERVIEW / RAUKAWA FM / CONSOLE
Douglas Bagnall
Raukawa FM
Tokoroa
MĀORI CONTENT:
8:35
12 September 2018
10.5 hours achieved
Te Reo 4:23
English speech 0:06
Māori music 0:59
Missing data 9:00
0:00 1:00 2:00 3:00 4:00 5:00 6:00 7:00 8:00 9:00 10:00 11:00 12:00 13:00 14:00 15:00 16:00 17:00 18:00 19:00 20:00 21:00 22:00 23:00 24:00
Today 12
Tue Sep 11
Mon Sep 10
Sun Sep 09
Sat Sep 08
Fri Sep 07
Thu Sep 06
Wed Sep 05
Tue Sep 04
Mon Sep 03
Sun Sep 02
Sat Sep 01

https://github.com/douglasbagnall/recur

douglas.bagnall@catalyst.net.nz