Switched Attribute and Spatial Coordinates
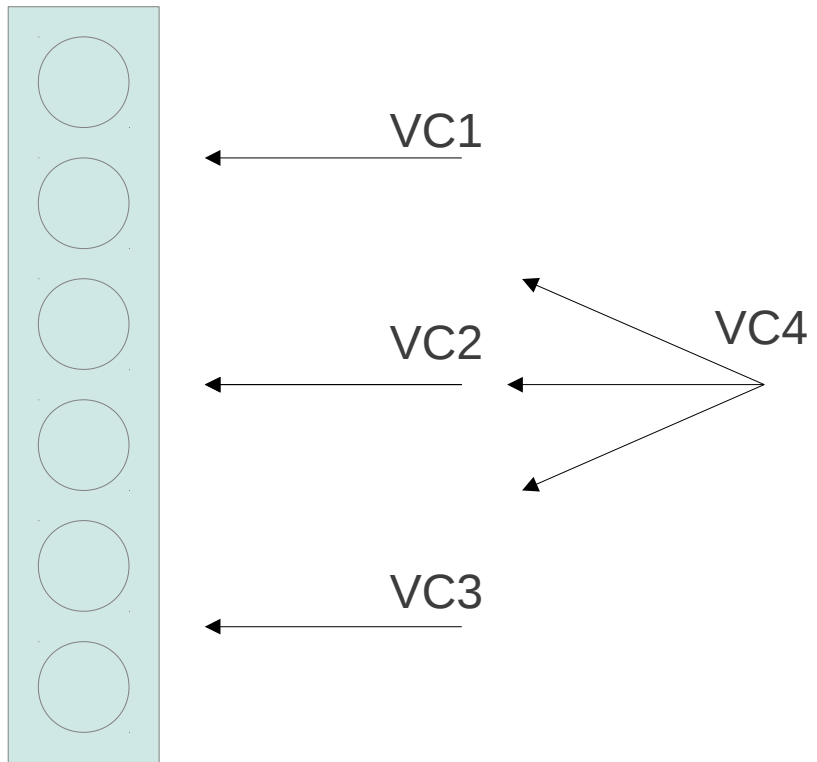
IETF CLUE Interim Meeting
June 2012

Andy Pepperell
apeppere@gmail.com

Switched attribute usage

- Usual case, 6 seats / 3 cameras:

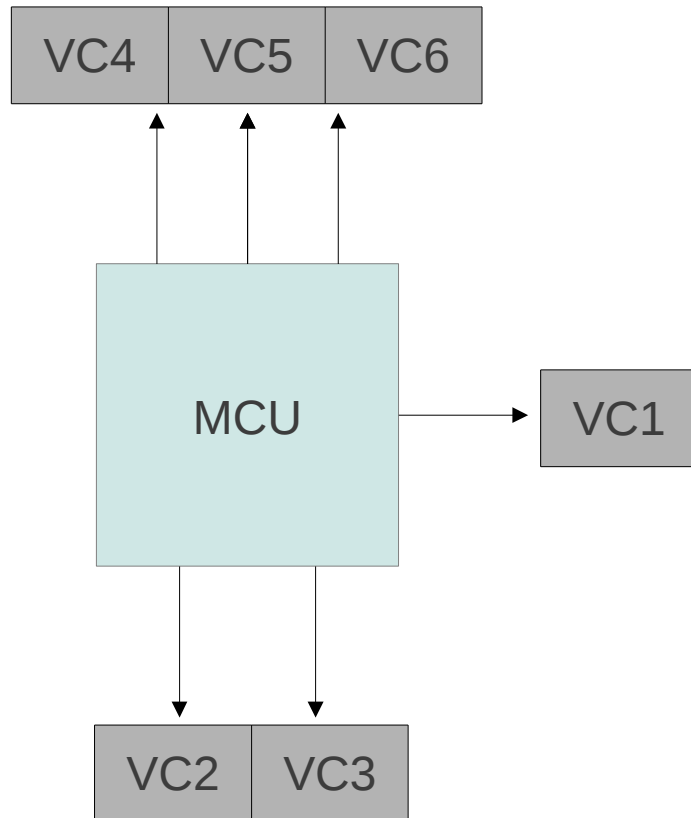| Capture scene entries	
1	VC1, VC2, VC3
2	VC4 ("switched")
3	*[VC5 zoomed out view]*

VC1

VC2

VC4

VC3

("switched" attribute may be the only way to distinguish VC4 and VC5 zoomed out view)

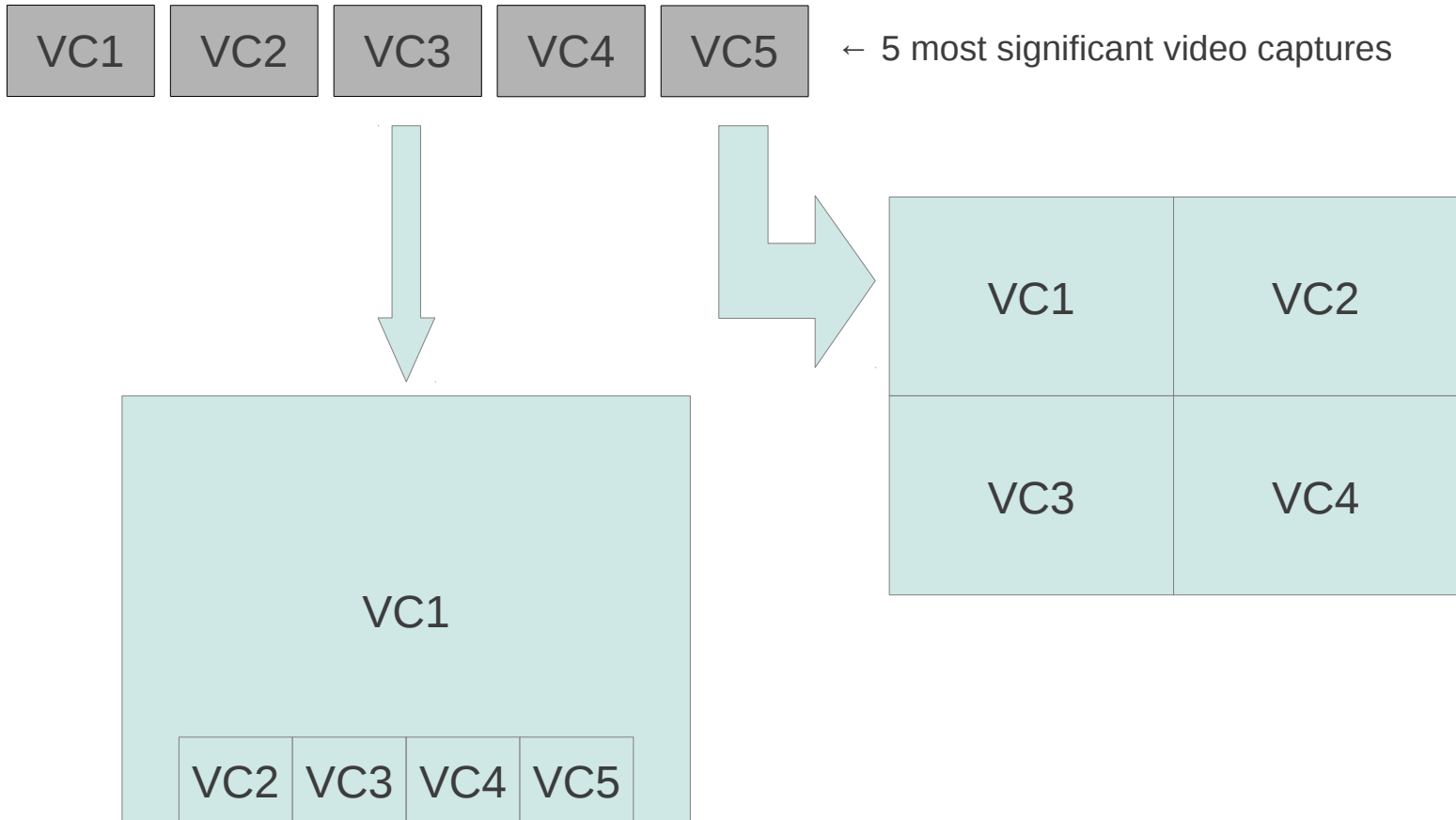# MCU (transcoded) case

- Handle 1, 2, and 3 screen consumers

VC4	VC5	VC6

MCU → VC1

VC2	VC3

Capture scene entries	
1	VC1
2	VC2, VC3
3	VC4, VC5, VC6

Provider-side co-ordinates for VC1 – VC6 allow correct rendering by consumer

What do we really want?

- Might be nice to allow:

| VC1 | VC2 | VC3 | VC4 | VC5 | ← 5 most significant video captures |

VC1	VC2
VC3	VC4

VC1

| VC2 | VC3 | VC4 | VC5 |

Previous slide explanation

- Middle box maintains ordered list of captures (VC1 – VC5)
  - typically ordered by audio energy
- Consumer performs own composition
- Desirable or essential…?
  - MCU may have no transcoding capabilities
  - Wish to reduce number of transcodes for quality or latency
  - Better / more appropriate layouts by consumer

Challenges

- Challenge 1:
  - Mechanism for advertisement and selection of ordered capture list
- Challenge 2:
  - Correct rendering of multi-capture sources from ordered capture list
- Challenge 3:
  - Lip sync of dynamically changing ordered captures

Challenge 1

- Different consumers want to use different number of captures, but undesirable to advertise full set:

Capture scene entries	
1	VC1
2	VC2, VC3
3	*VC4, VC5, VC6*
4	VC7, VC8, VC9, VC10
5	VC11, VC12, VC13, VC14, VC15
n	*"n" switched video captures*

(and similar for audio captures)

Challenge 1 proposal

- Proposed solution:
  - Use "switched" attribute and absence of provider-side capture co-ordinates to indicate a dynamic mapping
    - also implies validity of partial selection
  - differentiation of VCn and ACn needed for "transcoded" 1, 2, 3, 4, … screen cases but not ordered switched captures

Challenge 2

- Not splitting multi-capture sources; consider:

VC1	VC2	VC3

Ordered participant 1 ("loudest")

VC4

Ordered participant 2 ("second loudest")

VC5

Ordered participant 3
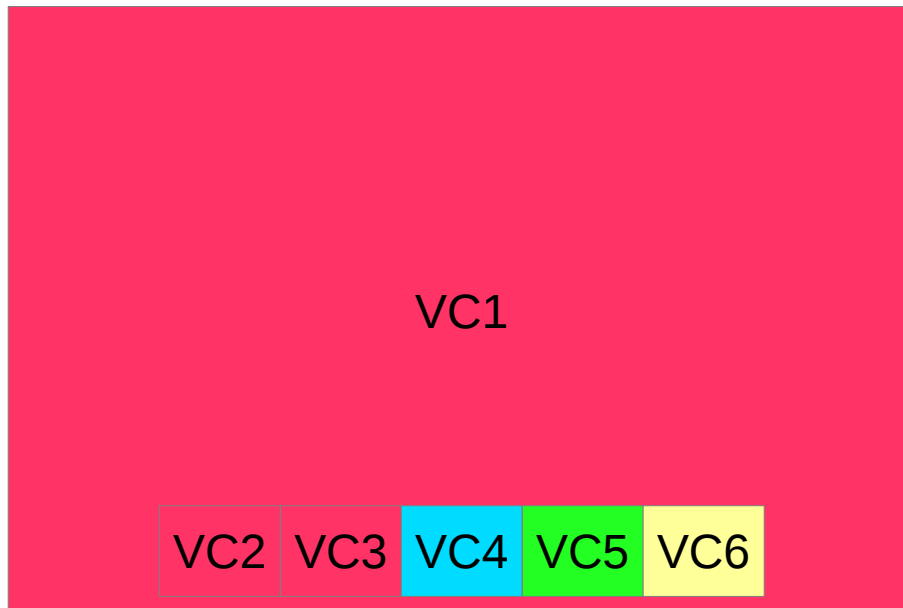
VC6

Ordered participant 4

Challenge 2

- Valid distribution across 3 screens:
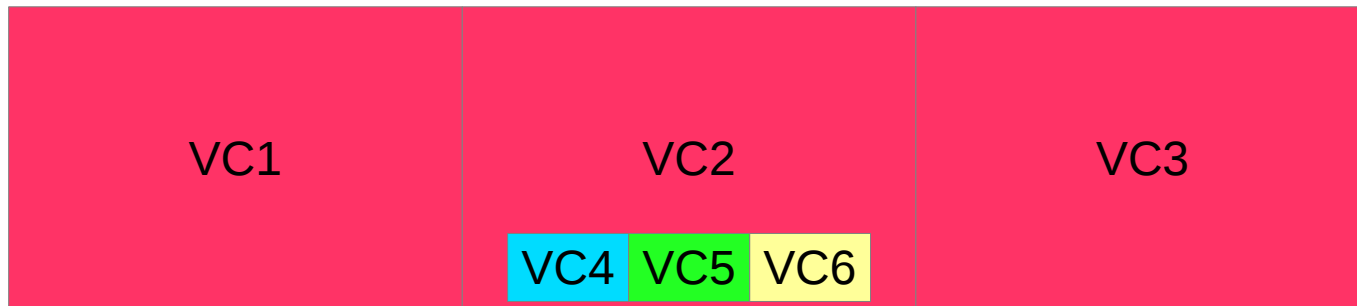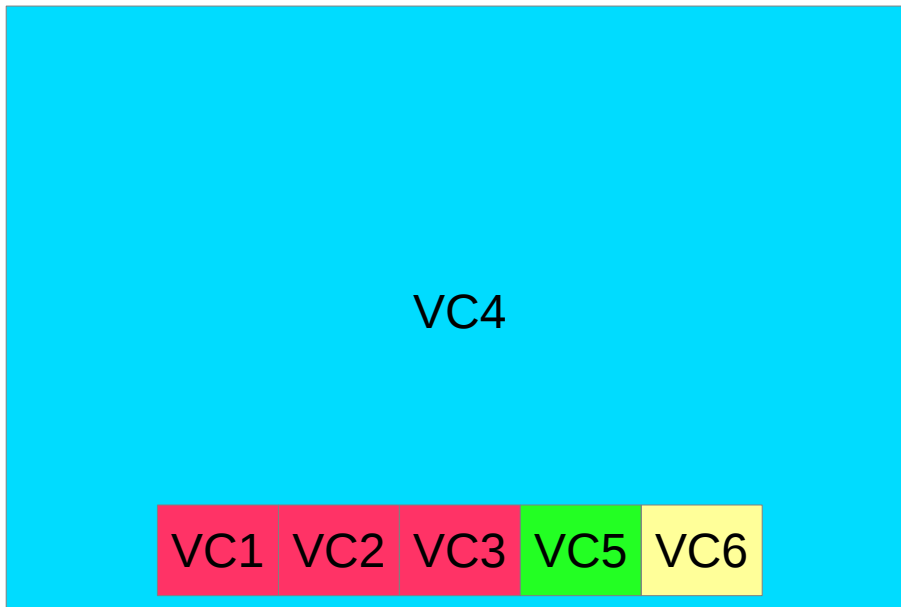
Challenge 2

- Naive, invalid, distribution of captures on 1 screen:

(VC1 – VC3 split incorrectly)

Challenge 2

- Valid distribution on 1 screen:

(VC1 – VC3 no longer split incorrectly)

Challenge 2 proposal

- Consumer needs to send some layout information to provider
  - Not full layout, but abstracted form of it
    - "rows" and "strips" simplifications considered
    - proposal instead is to introduce consumer-side render X, Y, Z co-ordinates
      - sent to provider along with consumer's capture instantiations
      - analog of provider co-ordinate scheme
      - dovetails co-ordinate absence for ordered capture list

Challenge 3

- With ordered audio and video captures changing frequently, how can lip sync be assured / achieved?
  - No static mapping between "n" ordered video captures and "m" ordered audio captures
    - Easy to see pathological cases when audio-only participants are involved, or when n != m

Challenge 3 proposal

- Ensure RTCP CNAME values used correctly
  - Should enable audio / video capture correspondence to be tracked
  - Expectation / assumption that active speaker order could change at a frequency that would require an RTP / RTCP solution here, i.e. that updates via CLUE messages would not be sufficiently "real time"

Alternative scheme

- Aforementioned proposals deliberately minimal
  - Absence of provider-side co-ordinates for "switched" captures implies capture scene entry from which it is valid to take a subset
    - Further implication that consumer should provider render-side co-ordinates
- Less minimal way of working:
  - Use an explicit capture set entry attribute to indicate this
  - However, implied case has the advantage of ascribing meaning to co-ordinate absence

Further study

- MCU ↔ MCU cascading cases; neither side providing co-ordinates
  - Consider EP → MCU → MCU → EP