Extraction of Read Text Using a Wearable Eye Tracker for Automatic Video Annotation

Mizuki Matsubara* Joachim Folz**
Takumi Toyama** Marcus Liwicki**
Andreas Dengel** Koichi Kise*

*Graduate school of Engineering Osaka Prefecture University
** German Research Center for Artificial Intelligence

Outline

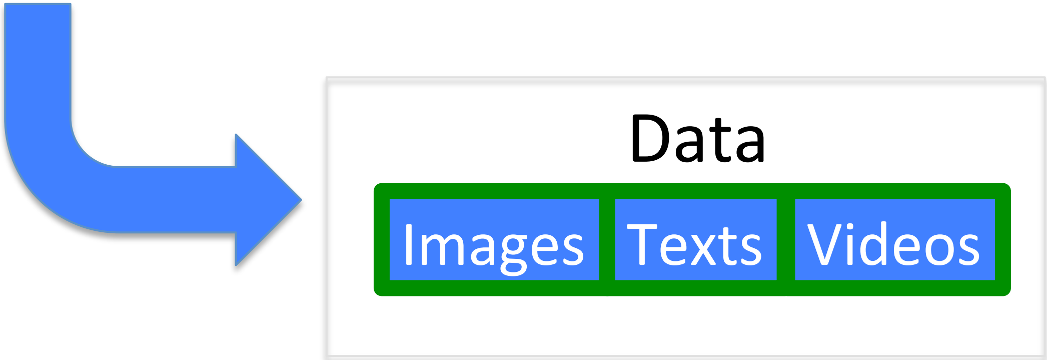
- Motivation
- Approach
- Pilot Study
- Conclusion & Future Work

Outline

- Motivation
- Approach
- Pilot Study
- Conclusion & Future Work

Life-logging

- Recording a person’s life

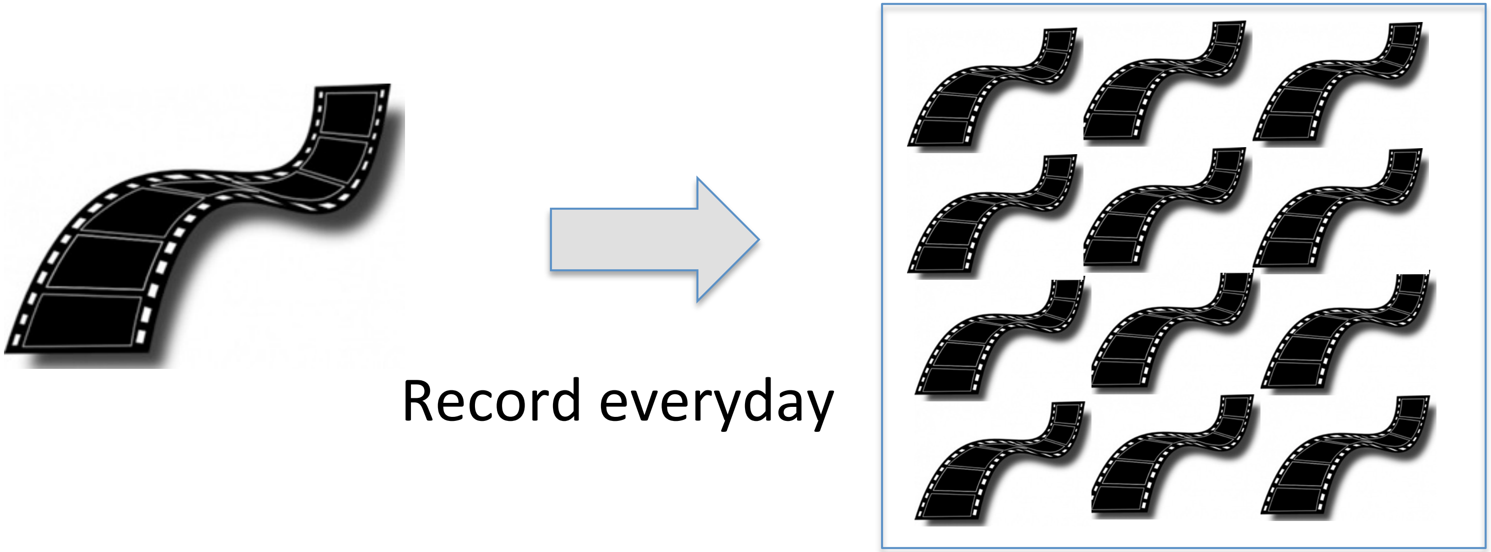


Data

Images Texts Videos

Life log video

- Life log video tends to be very long



Record everyday

Video annotation is needed to index the videos

Automatic video annotation

Video annotations describe the video contents

Object recognition

Chair

Activity recognition

Playing soccer

Image recognition is a difficult computational problem

J. Thomason, S. Venugopalan, S. Guadarrama, K. Saenko, R. Mooney, "Integrating language and vision to generate natural language descriptions of videos in the wild." Proceedings of the 25th International Conference on Computational Linguistics (2014).

Proposed method

Our method annotates videos in a particular situation

A user is following a textual manual

Worker following a manual

Cook following a recipe

Texts in the manual can describe the user's actions.

└→ Annotation

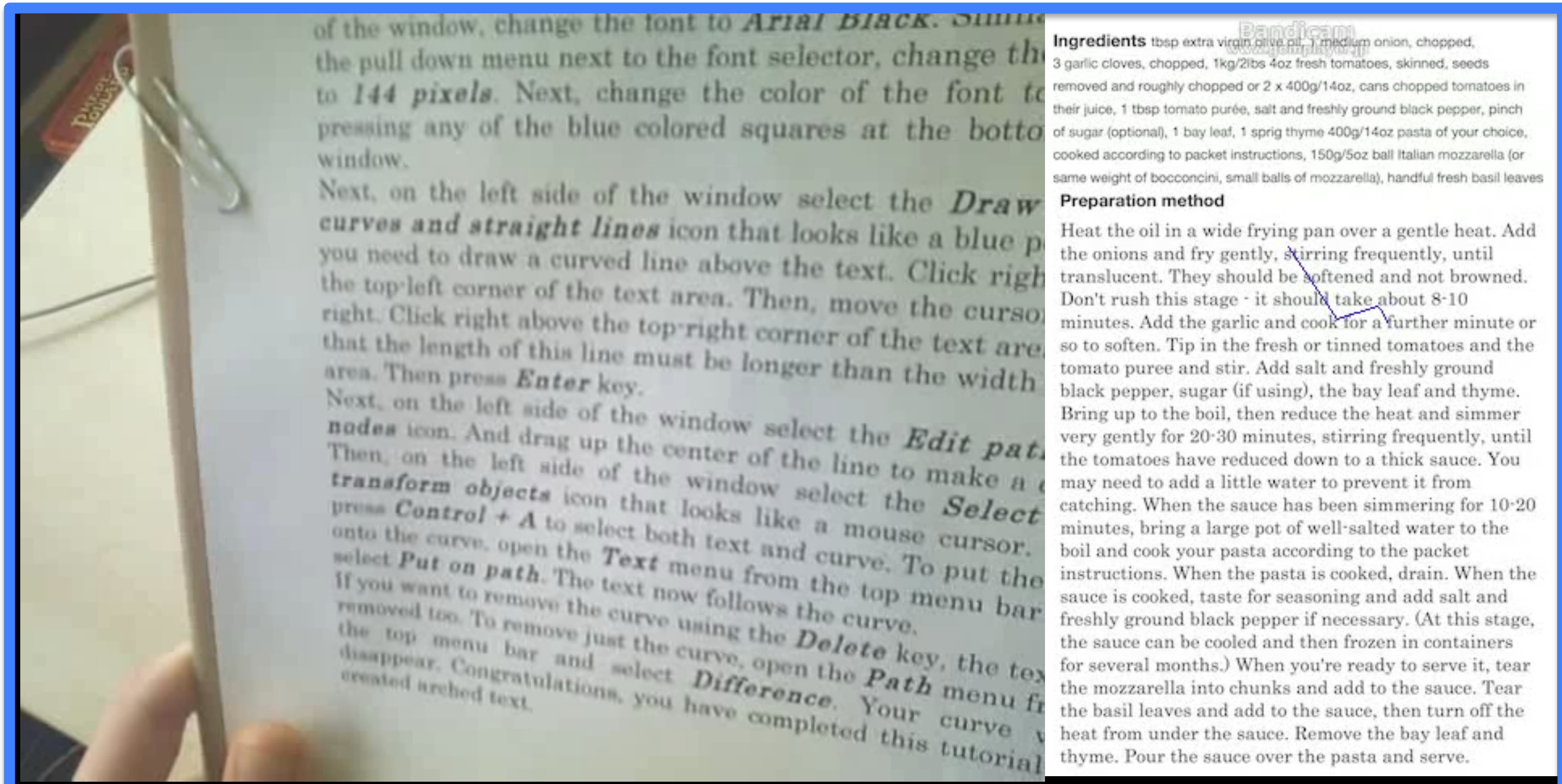
SMI's mobile eye tracker

- A device estimating the user's gaze

Camera

Record life log videos

Ideal system



of the window, change the font to *Arial Black*. Simi
the pull down menu next to the font selector, change th
to ***144 pixels***. Next, change the color of the font to
pressing any of the blue colored squares at the botto
window.
Next, on the left side of the window select the ***Draw***
curves and straight lines icon that looks like a blue p
you need to draw a curved line above the text. Click righ
the top-left corner of the text area. Then, move the curso
right. Click right above the top-right corner of the text are
that the length of this line must be longer than the width
area. Then press ***Enter*** key.
Next, on the left side of the window select the ***Edit pat***
nodes icon. And drag up the center of the line to make a c
Then, on the left side of the window select the ***Select***
transform objects icon that looks like a mouse cursor.
press ***Control + A*** to select both text and curve. To put the
onto the curve, open the ***Text*** menu from the top menu bar
select ***Put on path***. The text now follows the curve.
If you want to remove the curve using the ***Delete*** key, the tex
removed too. To remove just the curve, open the ***Path*** menu fr
the top menu bar and select ***Difference***. Your curve v
disappear. Congratulations, you have completed this tutorial
created arched text.

Ingredients tbsp extra virgin olive oil, 1 medium onion, chopped, 3 garlic cloves, chopped, 1kg/2lbs 4oz fresh tomatoes, skinned, seeds removed and roughly chopped or 2 x 400g/14oz, cans chopped tomatoes in their juice, 1 tbsp tomato purée, salt and freshly ground black pepper, pinch of sugar (optional), 1 bay leaf, 1 sprig thyme 400g/14oz pasta of your choice, cooked according to packet instructions, 150g/5oz ball Italian mozzarella (or same weight of bocconcini, small balls of mozzarella), handful fresh basil leaves

Preparation method

Heat the oil in a wide frying pan over a gentle heat. Add the onions and fry gently, stirring frequently, until translucent. They should be softened and not browned. Don't rush this stage - it should take about 8-10 minutes. Add the garlic and cook for a further minute or so to soften. Tip in the fresh or tinned tomatoes and the tomato puree and stir. Add salt and freshly ground black pepper, sugar (if using), the bay leaf and thyme. Bring up to the boil, then reduce the heat and simmer very gently for 20-30 minutes, stirring frequently, until the tomatoes have reduced down to a thick sauce. You may need to add a little water to prevent it from catching. When the sauce has been simmering for 10-20 minutes, bring a large pot of well-salted water to the boil and cook your pasta according to the packet instructions. When the pasta is cooked, drain. When the sauce is cooked, taste for seasoning and add salt and freshly ground black pepper if necessary. (At this stage, the sauce can be cooled and then frozen in containers for several months.) When you're ready to serve it, tear the mozzarella into chunks and add to the sauce. Tear the basil leaves and add to the sauce, then turn off the heat from under the sauce. Remove the bay leaf and thyme. Pour the sauce over the pasta and serve.

Draw a line above the text.

Example 1(following a manual)

When a worker doesn’t understand the manual

Video recording while working

Retrieve

Education of
a new employee

Watching the video help him

Example 2 (following a recipe)

When a cook forgets how to cook the meal he/she made before

Video recording while cooking

Retrieve

Aids of memory

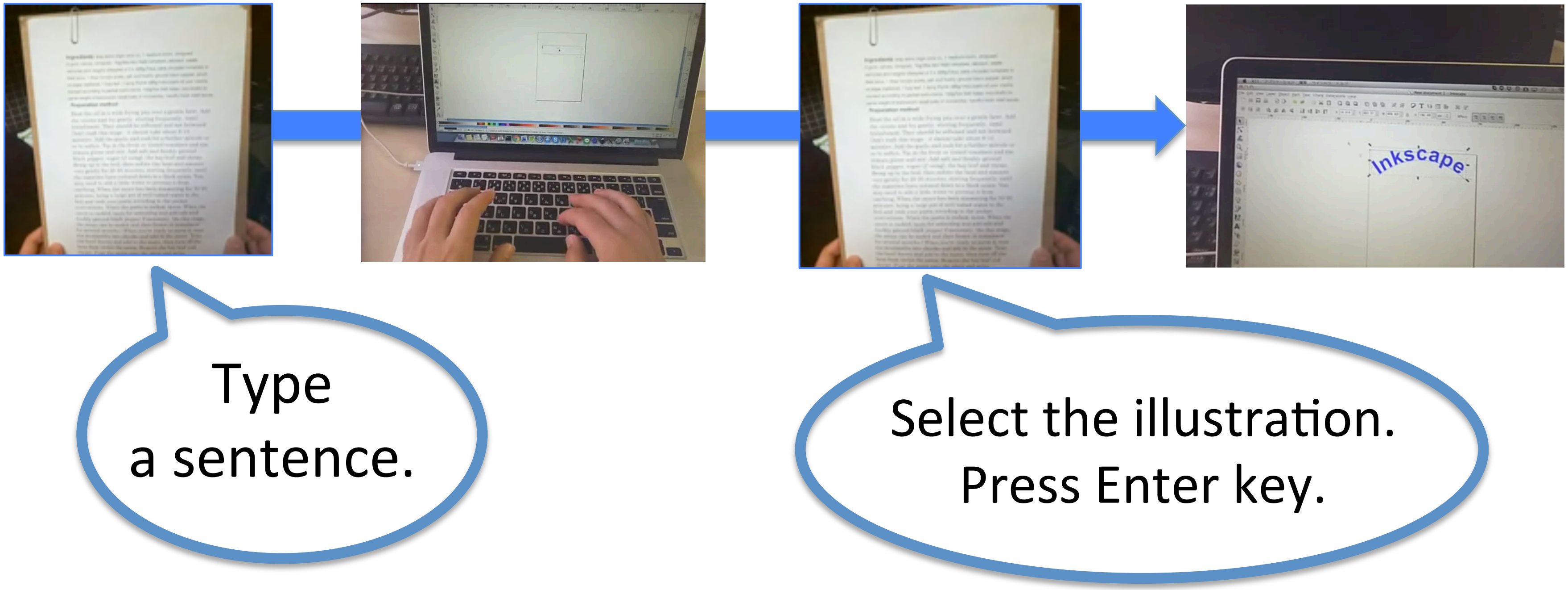
The video reminds him/her the way how to cook the meal

Outline

- Motivation
- Approach
- Pilot Study
- Conclusion & Future Work

Assumption

Users actions can be annotated with texts read just before



Type a sentence.

Select the illustration. Press Enter key.

Assumption

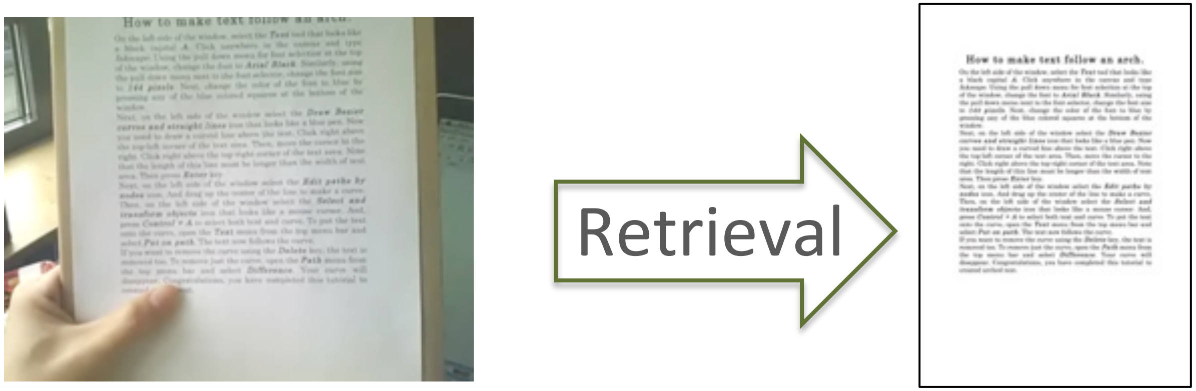
Document

Reading	Non-reading	Reading	Non-reading

Video

Retrieval of document image

- LLAH (Locally Likely Arrangement Hashing)



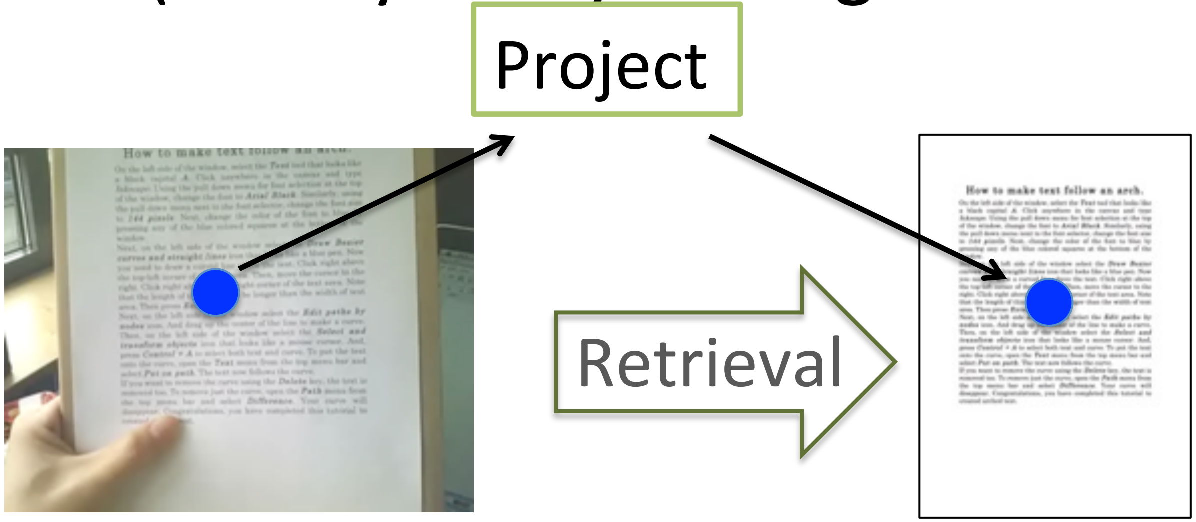
Video

Document image

Kai Kunze, Hitoshi Kawaichi, Kazuyo Yoshimura, and Koichi Kise, Towards inferring language expertise using eye tracking, CHI'13 Extended Abstracts on Human Factors in Computing Systems, 217–222, 2013.

Retrieval of document image

- LLAH (Locally Likely Arrangement Hashing)



Project

Retrieval

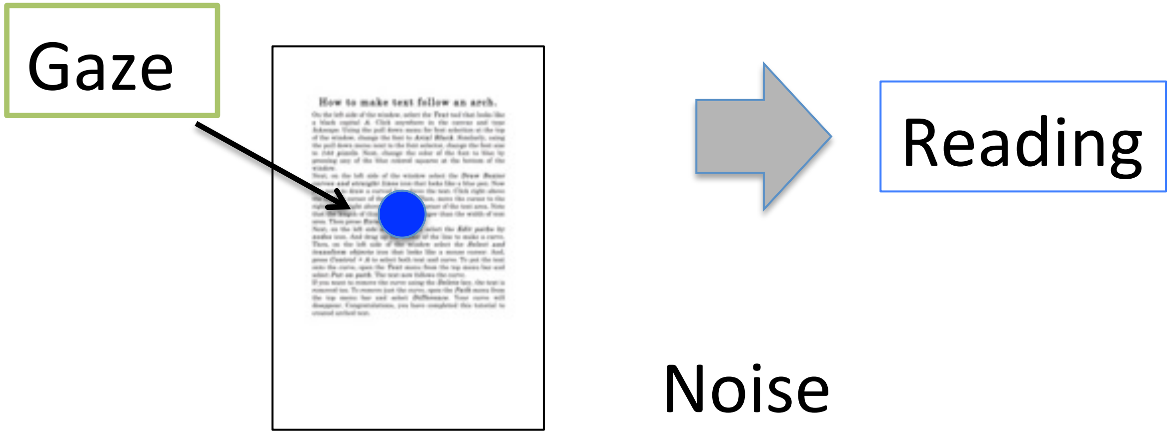
Video

Document image

Kai Kunze, Hitoshi Kawaichi, Kazuyo Yoshimura, and Koichi Kise, Towards inferring language expertise using eye tracking, CHI'13 Extended Abstracts on Human Factors in Computing Systems, 217–222, 2013.

Reading detection

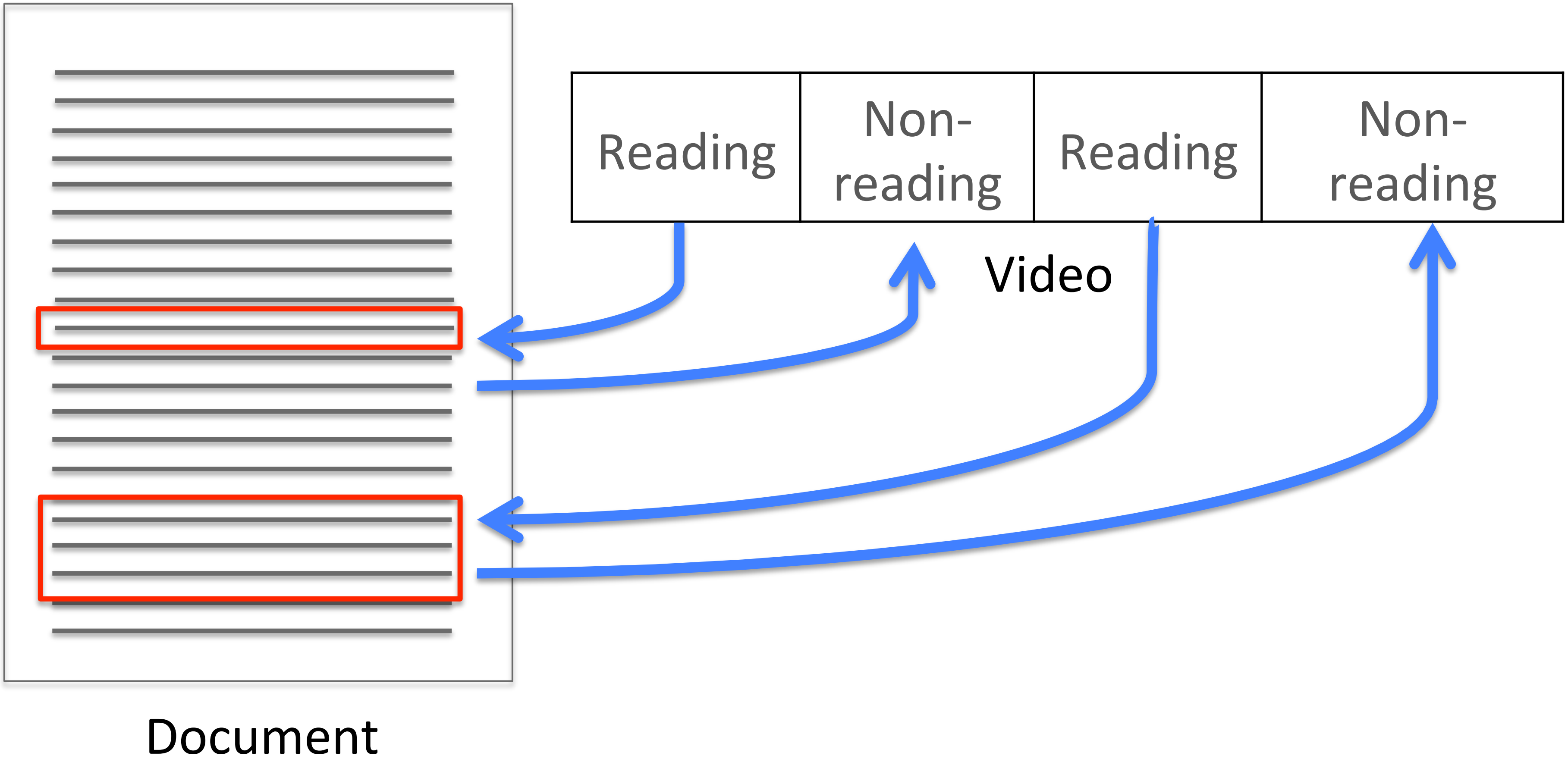
We assume that the user is reading the document while his/her gaze point is on the document



Noise

- influence of the user’s blink
- error of the eye tracker
- failure of retrieving the document image

Assumption

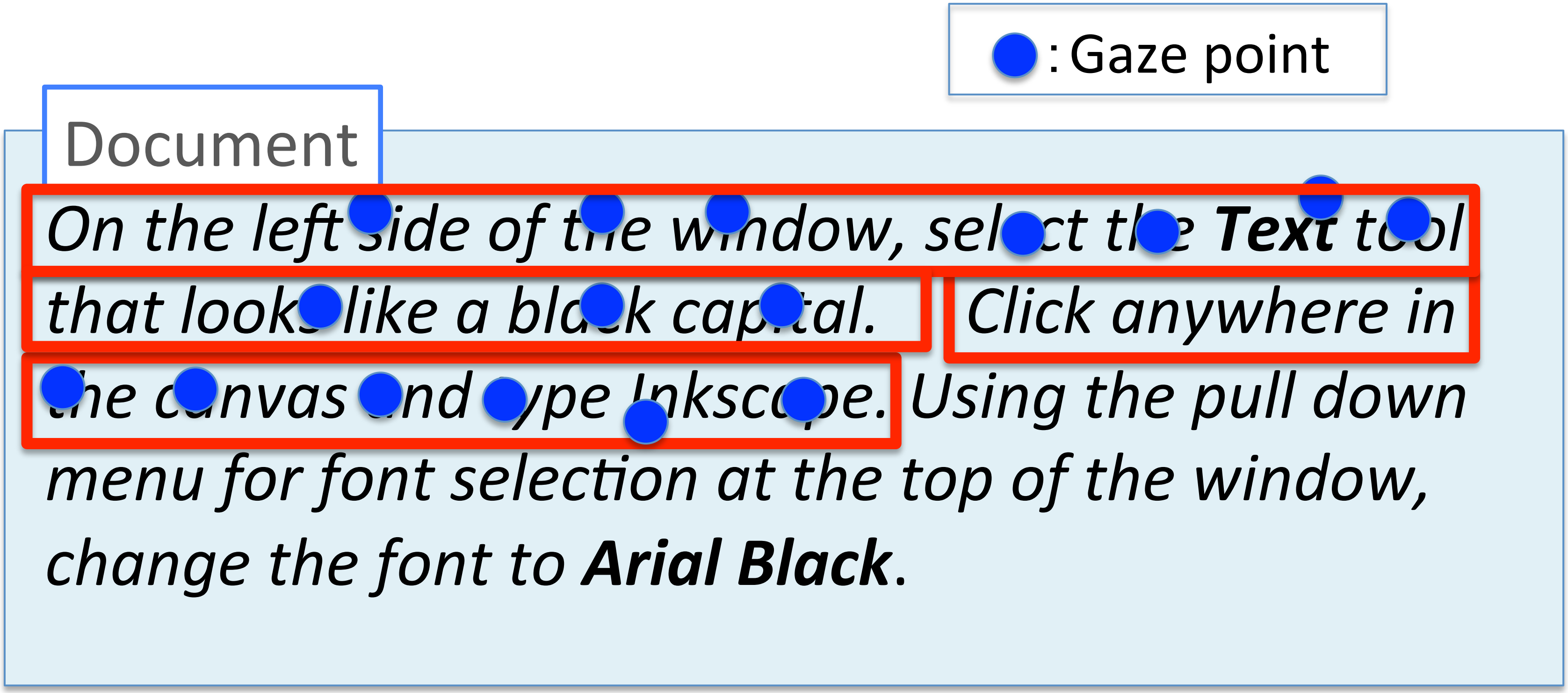


Reading | Non-reading | Reading | Non-reading

Video

Document

Extraction of read text



On the left side of the window, select the ***Text*** *tool that looks like a black capital. Click anywhere in the canvas and type Inkscape. Using the pull down menu for font selection at the top of the window, change the font to* ***Arial Black****.*

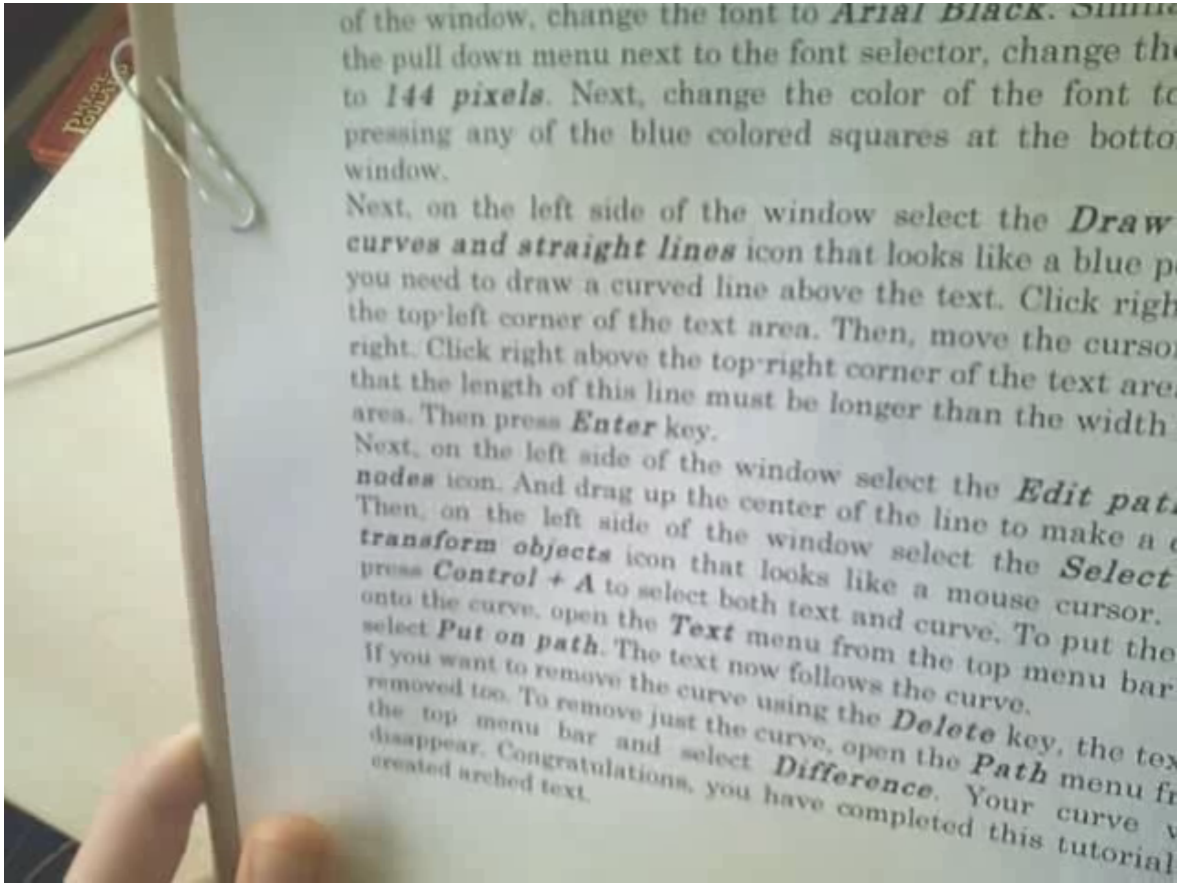
Extract the nearest sentences from the gaze points

Outline

- Motivation
- Approach
- Pilot Study
- Conclusion & Future Work

Task

- Follow a tutorial
  - Draw an illustration



of the window, change the font to ***Arial Black***. Simi
the pull down menu next to the font selector, change th
to ***144 pixels***. Next, change the color of the font to
pressing any of the blue colored squares at the botto
window.
Next, on the left side of the window select the ***Draw***
curves and straight lines icon that looks like a blue p
you need to draw a curved line above the text. Click righ
the top-left corner of the text area. Then, move the curso
right. Click right above the top-right corner of the text are
that the length of this line must be longer than the width
area. Then press ***Enter*** key.
Next, on the left side of the window select the ***Edit pat***
nodes icon. And drag up the center of the line to make a
Then, on the left side of the window select the ***Select***
transform objects icon that looks like a mouse cursor.
press ***Control + A*** to select both text and curve. To put the
onto the curve, open the ***Text*** menu from the top menu bar
select ***Put on path***. The text now follows the curve.
If you want to remove the curve using the ***Delete*** key, the tex
removed too. To remove just the curve, open the ***Path*** menu fr
the top menu bar and select ***Difference***. Your curve
disappear. Congratulations, you have completed this tutorial
created arched text.

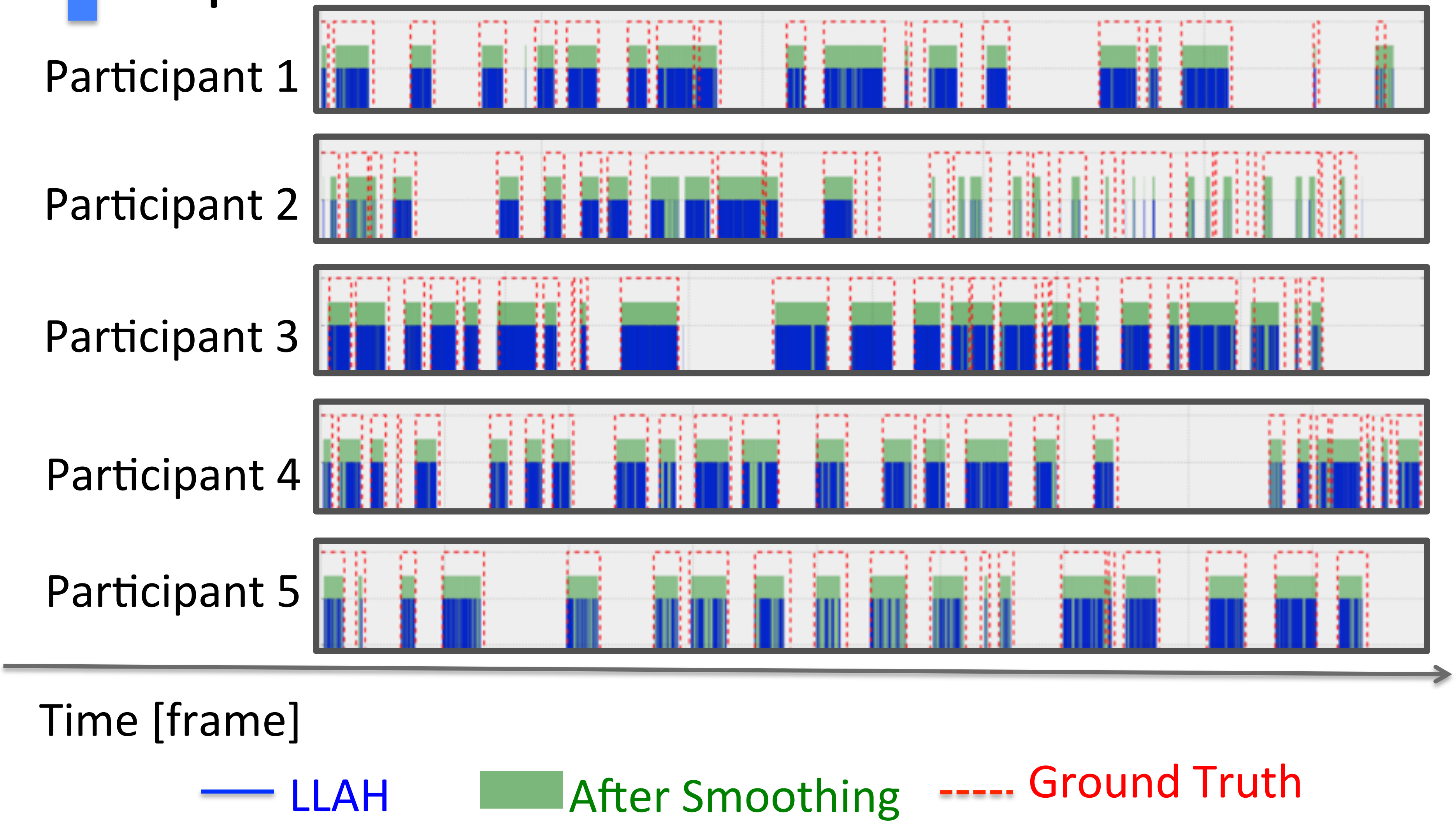
Experimental condition

- Participant : 5 (20～30’s males and females)
- Recording time : 345, 458, 393, 370, 354 [seconds]
- Language : English
- One page

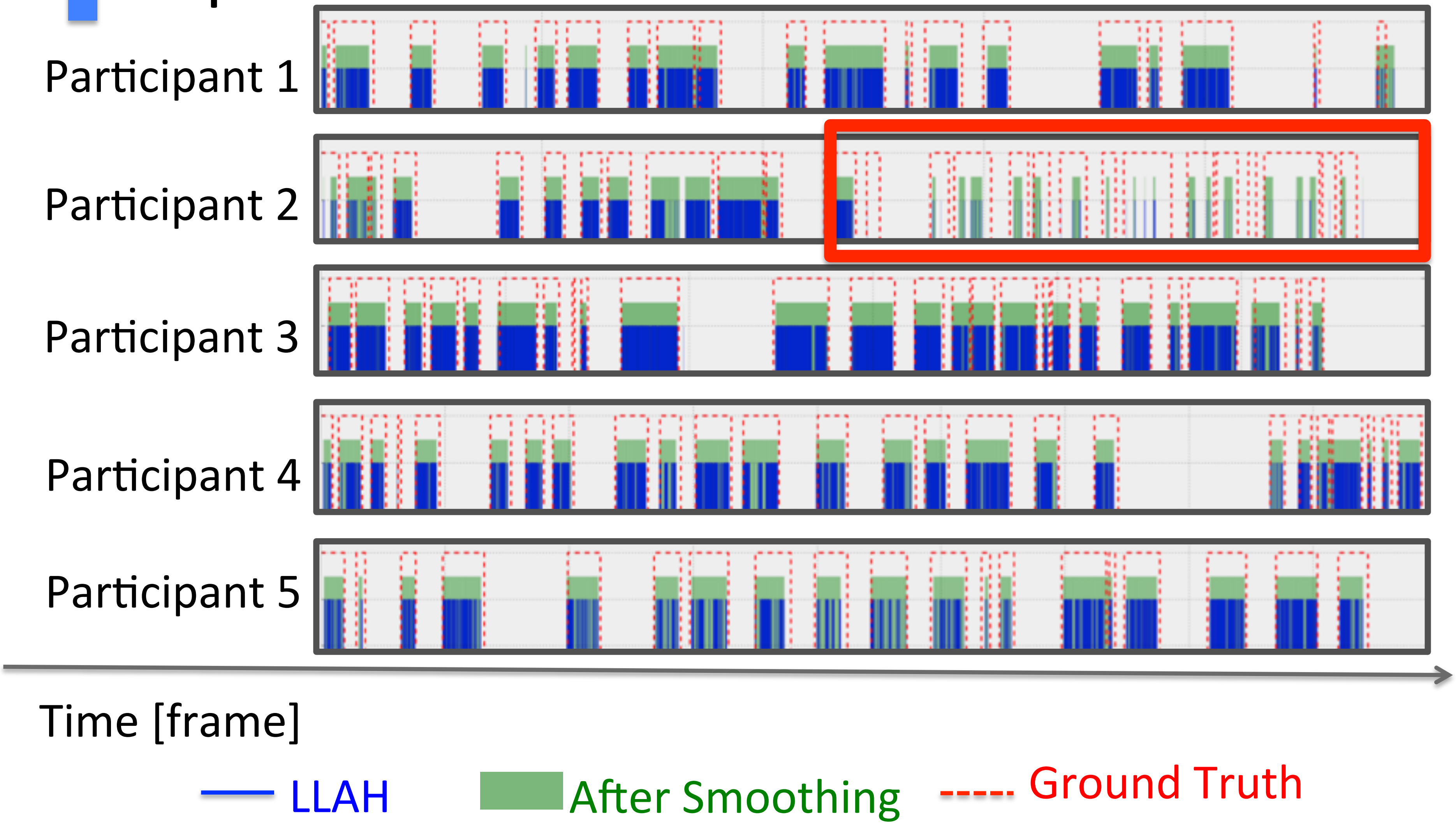
Evaluation of the reading detection

- We applied the reading detection to the recording videos.
  - examine the accuracy of the reading detection

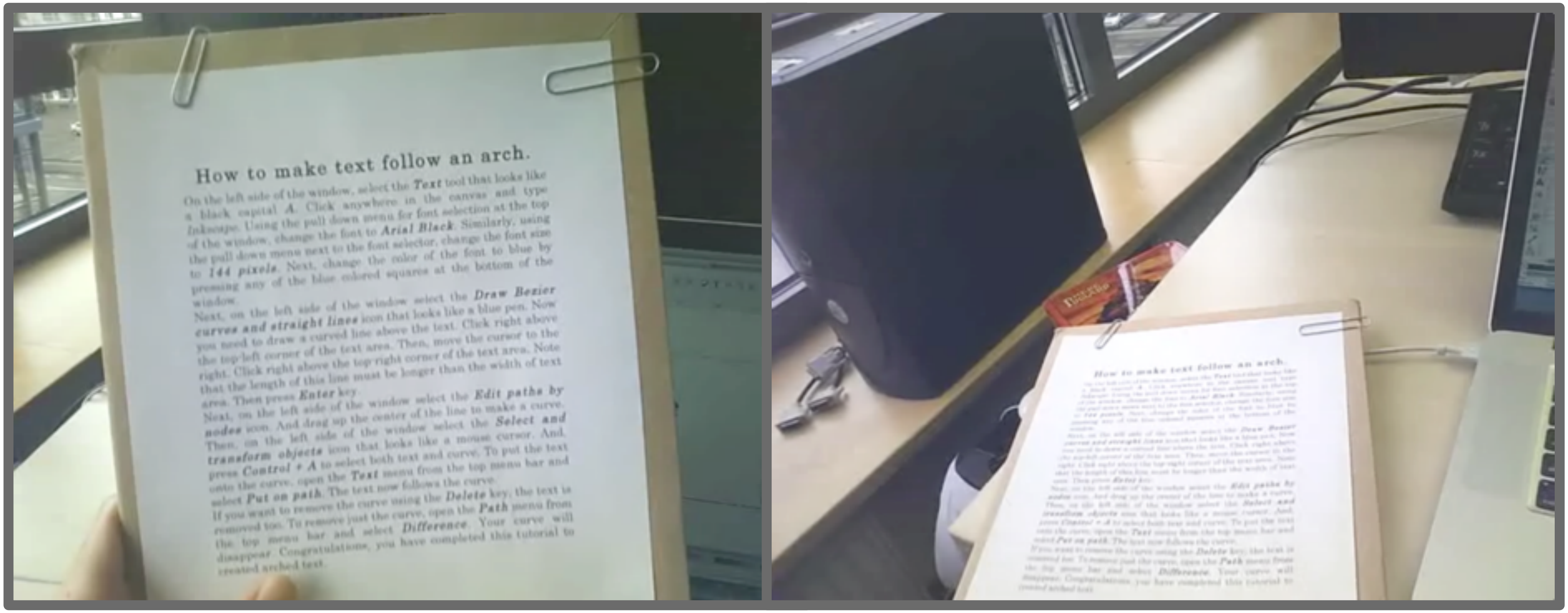
Experimental result



Experimental result



Examples of recording frames



Success

Failure

How to make text follow an arch.

On the left side of the window, select the ***Text*** tool that looks like a black capital *A*. Click anywhere in the canvas and type Inkscape. Using the pull down menu for font selection at the top of the window, change the font to ***Arial Black***. Similarly, using the pull down menu next to the font selector, change the font size to ***144 pixels***. Next, change the color of the font to blue by pressing any of the blue colored squares at the bottom of the window.

Next, on the left side of the window select the ***Draw Bezier curves and straight lines*** icon that looks like a blue pen. Now you need to draw a curved line above the text. Click right above the top-left corner of the text area. Then, move the cursor to the right. Click right above the top-right corner of the text area. Note that the length of this line must be longer than the width of text area. Then press ***Enter*** key.

Next, on the left side of the window select the ***Edit paths by nodes*** icon. And drag up the center of the line to make a curve.

Then, on the left side of the window select the ***Select and transform objects*** icon that looks like a mouse cursor. And, press ***Control + A*** to select both text and curve. To put the text onto the curve, open the ***Text*** menu from the top menu bar and select ***Put on path***. The text now follows the curve.

If you want to remove the curve using the ***Delete*** key, the text is removed too. To remove just the curve, open the ***Path*** menu from the top menu bar and select ***Difference***. Your curve will disappear. Congratulations, you have completed this tutorial to created arched text.

The distance and angle influence the retrieval of LLAH

Accuracy of the reading detection

Recall

$$\frac{\text{successfully detected reading frames}}{\text{Ground truth frames}}$$

$\fallingdotseq$ 80.0%

Precision

$$\frac{\text{successfully detected reading frames}}{\text{Retrieved document frames}}$$

$\fallingdotseq$ 100%

Pilot study

Assumption

Users actions can be annotated with texts read just before

Purpose

Examine how our assumption works while participants are following a manual naturally

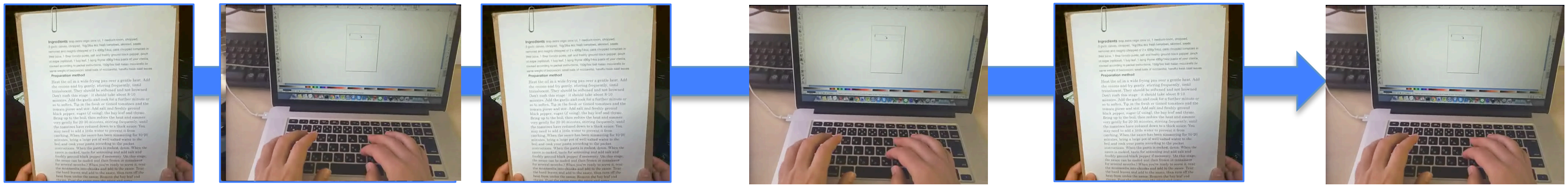
Experimental condition

- Participant : 5 (20～30's males and females)
- Recording time : 345, 458, 393, 370, 354 [seconds]
- Language : English
- One page
- Rule out video segmentation error

Correct the video segmentation

Correct to “Reading”

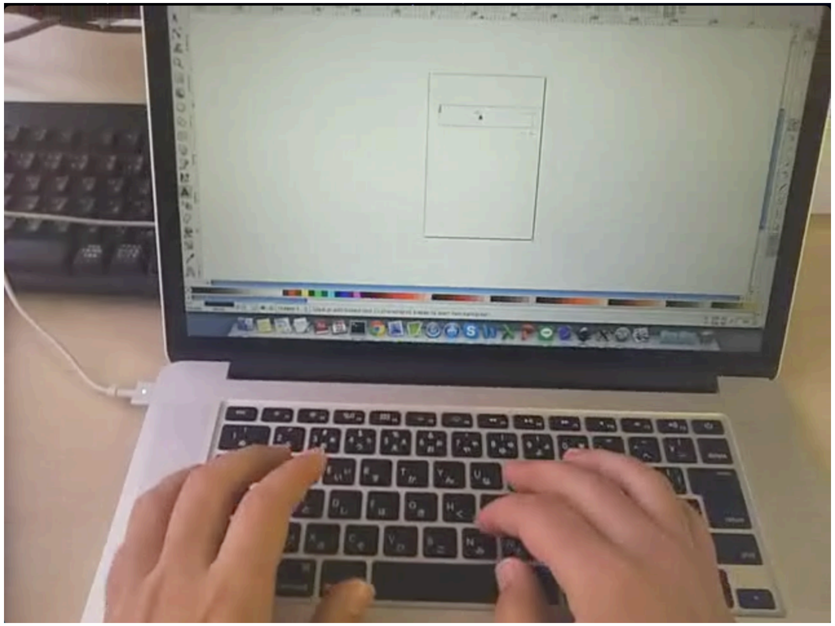
Reading	Not reading	Not reading	Not reading	Reading	Not reading

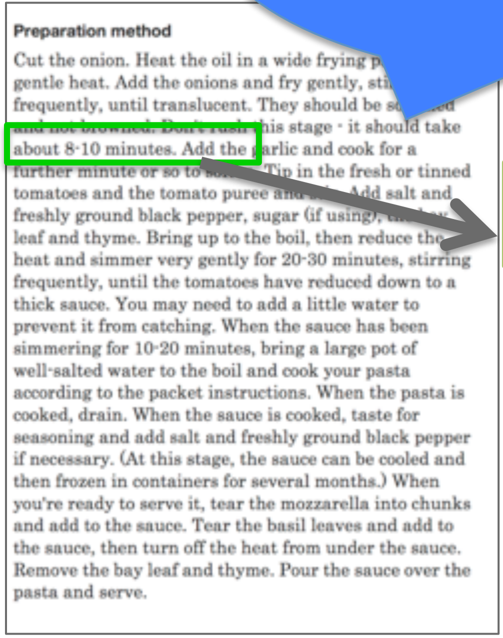


Evaluation of the extraction of the annotations

Compared the annotations to the correct annotations

Correct annotation: manually annotate





Preparation method

Cut the onion. Heat the oil in a wide frying p gentle heat. Add the onions and fry gently, sti frequently, until translucent. They should be so ed and not browned. Don't rush this stage - it should take about 8-10 minutes. Add the garlic and cook for a further minute or so to so Tip in the fresh or tinned tomatoes and the tomato puree and Add salt and freshly ground black pepper, sugar (if using), the bay leaf and thyme. Bring up to the boil, then reduce the heat and simmer very gently for 20-30 minutes, stirring frequently, until the tomatoes have reduced down to a thick sauce. You may need to add a little water to prevent it from catching. When the sauce has been simmering for 10-20 minutes, bring a large pot of well-salted water to the boil and cook your pasta according to the packet instructions. When the pasta is cooked, drain. When the sauce is cooked, taste for seasoning and add salt and freshly ground black pepper if necessary. (At this stage, the sauce can be cooled and then frozen in containers for several months.) When you're ready to serve it, tear the mozzarella into chunks and add to the sauce. Tear the basil leaves and add to the sauce, then turn off the heat from under the sauce. Remove the bay leaf and thyme. Pour the sauce over the pasta and serve.

Type a sentence.

Experimental result

Recall[%]	Precision[%]
64.5	30.8

Recall:

Actions which can be annotated by the read texts are correctly annotated at the rate of 64.5%.

Precision:

Annotations are correct at the rate of 30.8%.

Discussion

Causes for error:

1. Extraction of read sentences fails

   If we could extract read texts correctly,

   Recall : 86.8 %

   Precision : 61.0%

2. Assumption sometimes does not work

Outline

- Motivation
- Approach
- Pilot Study
- Conclusion & Future Work

Conclusion & Future work

Conclusion

- Proposed a method for automatic video annotation in the scenario where users are following a textual manual
- Accuracy :

  If the video segmentation is succeeded,

  – Recall : 64.5%   Precision : 30.8%

Future Work

- Improve the recall and precision of the annotation
- Improve the video segmentation method