Some visualization result about speech engrave (text attend speech)

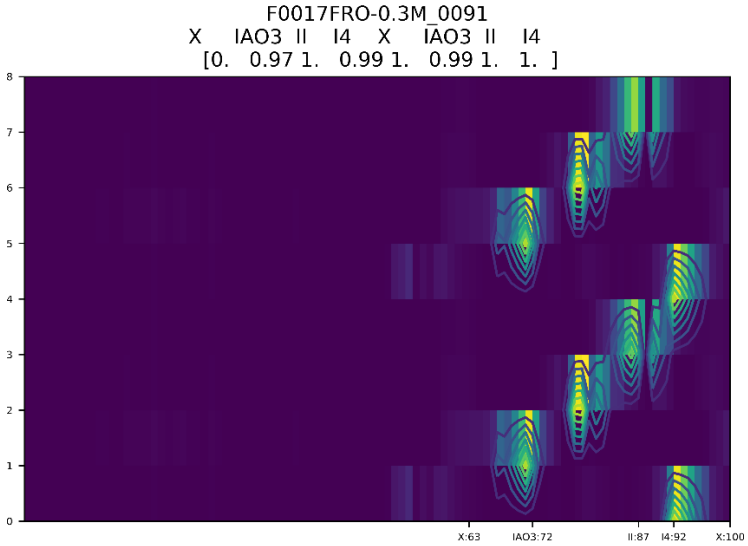
The first group is the Recall Group.

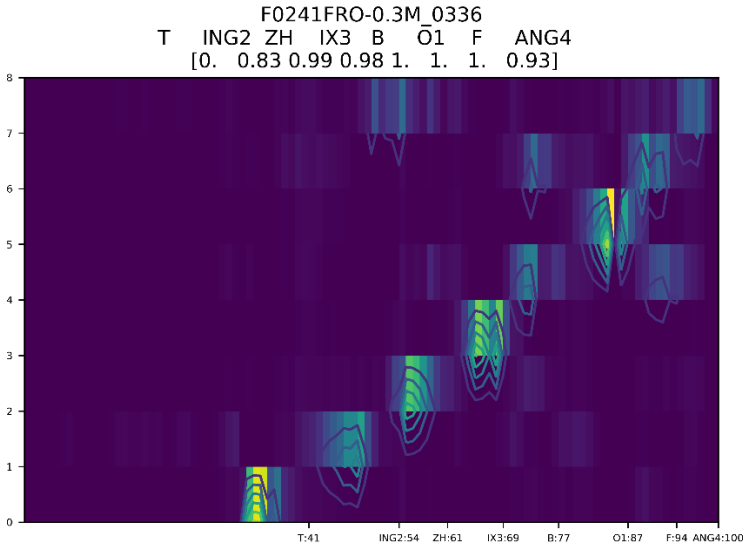
The title of the figure is: [utterance / keywords phone sequence / score of each phone]

Y axis: keywords index

X axis: time line with label (pronounced phone and there index e.g. X:20 means phone “x” end at 19th frame.)

Recall attention weight visualization.



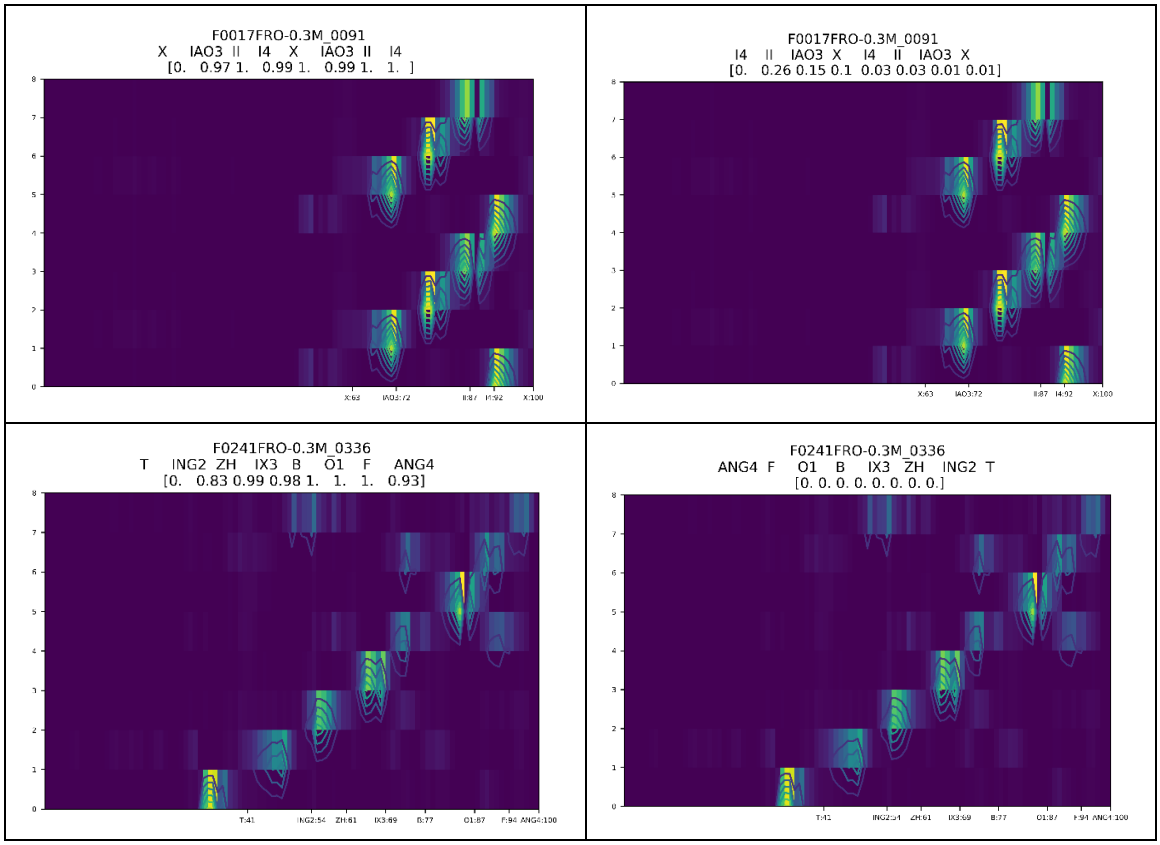


After that we reverse the keywords phone sequence e.g:

A B C D => D C B A

The attention weight is shown as follow:

The left column is the original attention weight and the right column is the reversed attention weight.

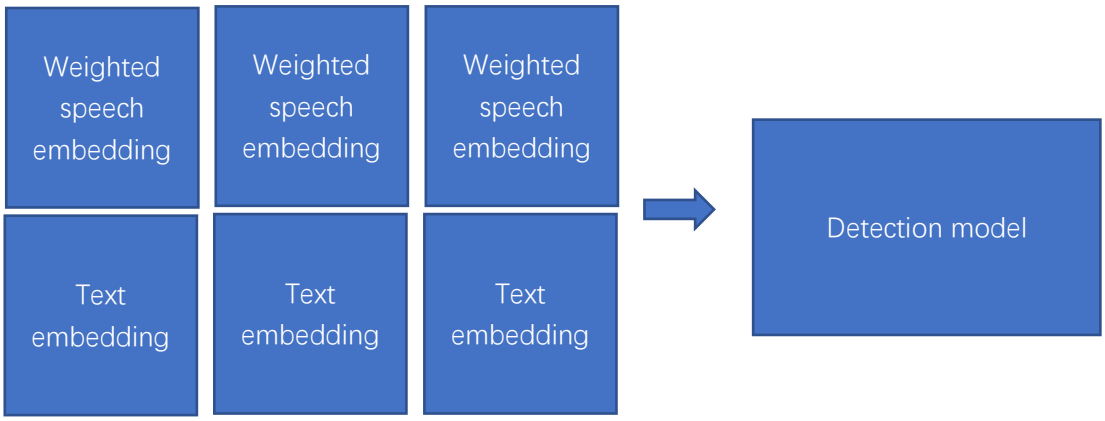


Intuitively, even we reverse the input sequence the attention weight shouldn't be changed. (In the figure we re-range the attention weight of right column)

We should note that even the attention weight is not change. But the

decreasing of the score is remarkable.

Most of them are around zero.



As shown above the input of detection model is

{ [text embedding: weighted speech embedding]$_1$; [text embedding: weighted speech embedding]$_2$ …}

So even we change the Text embedding order. The input pair still matched.

So the model should answer YES. But most of them are around zero.

There are tow possible reasons:

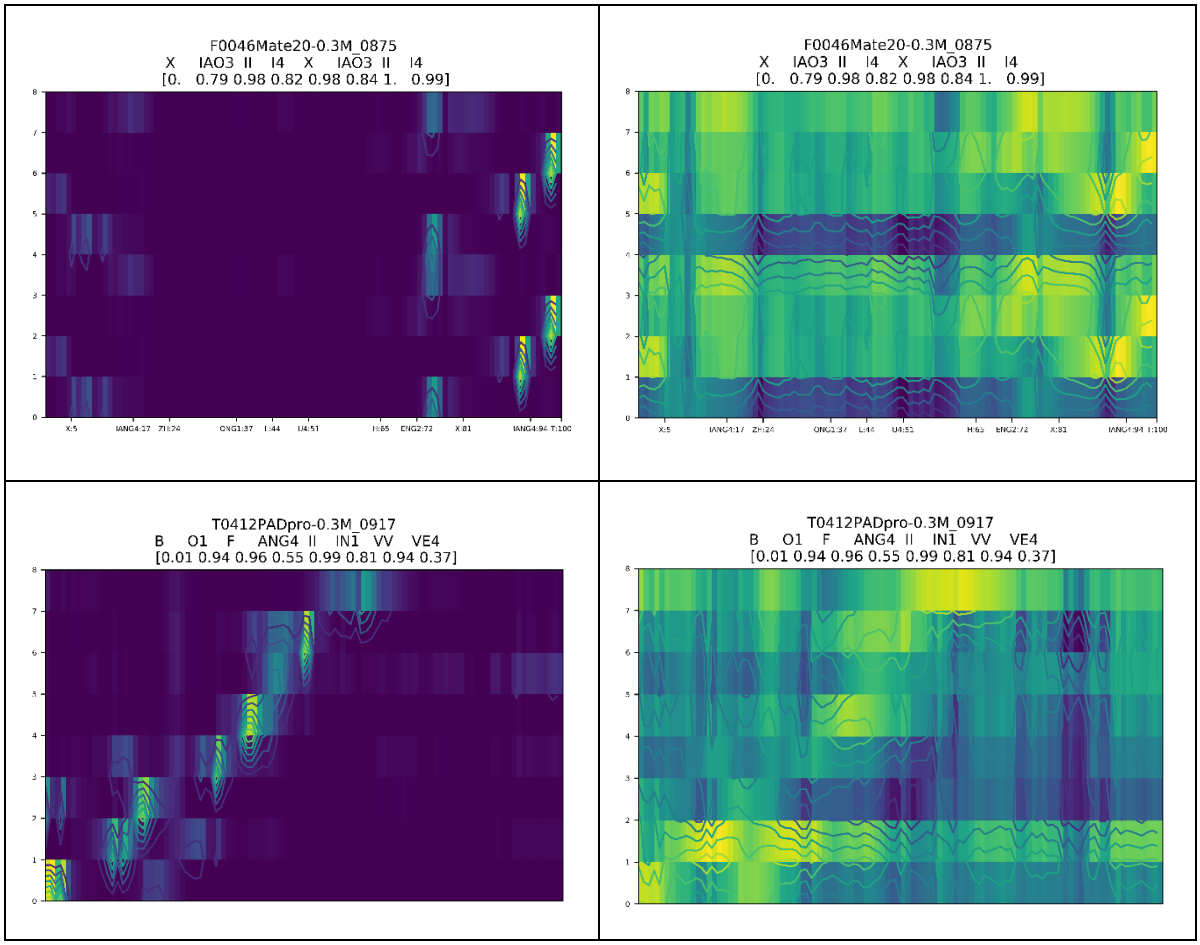
1. Only negative samples in training data are sorted by un-normal order.
2. The RNN of detection model learn some information about LM. But we hope that detection model should only care about whether the input pair is matched or not.

The Impact of softmax function which is employed by attention

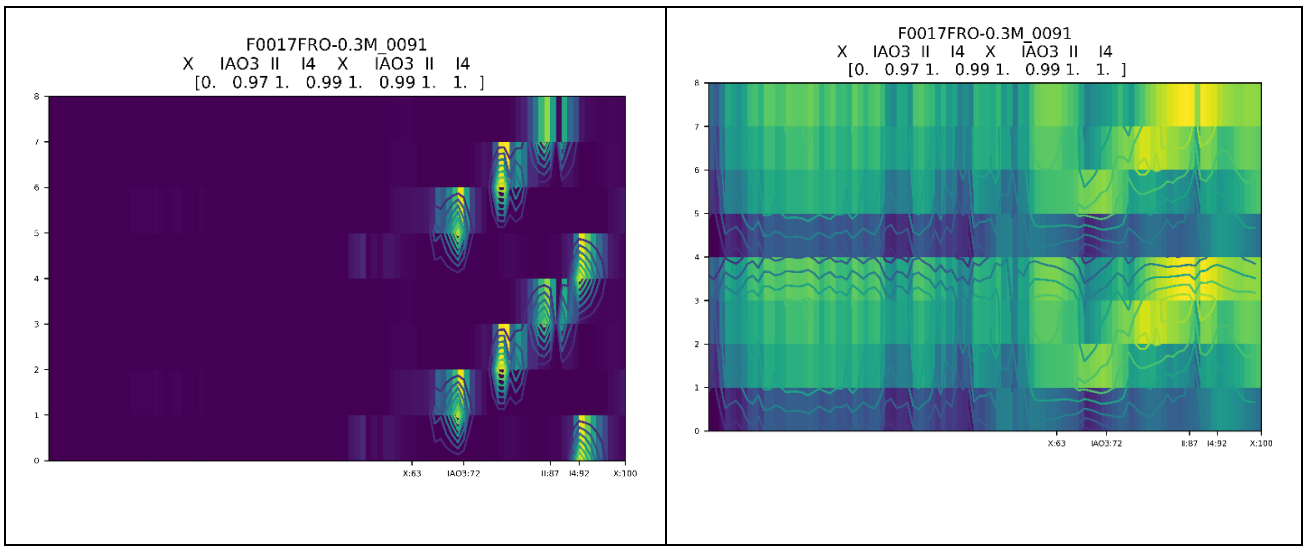
The left column is the original attention weight(re-scaled by softmax)

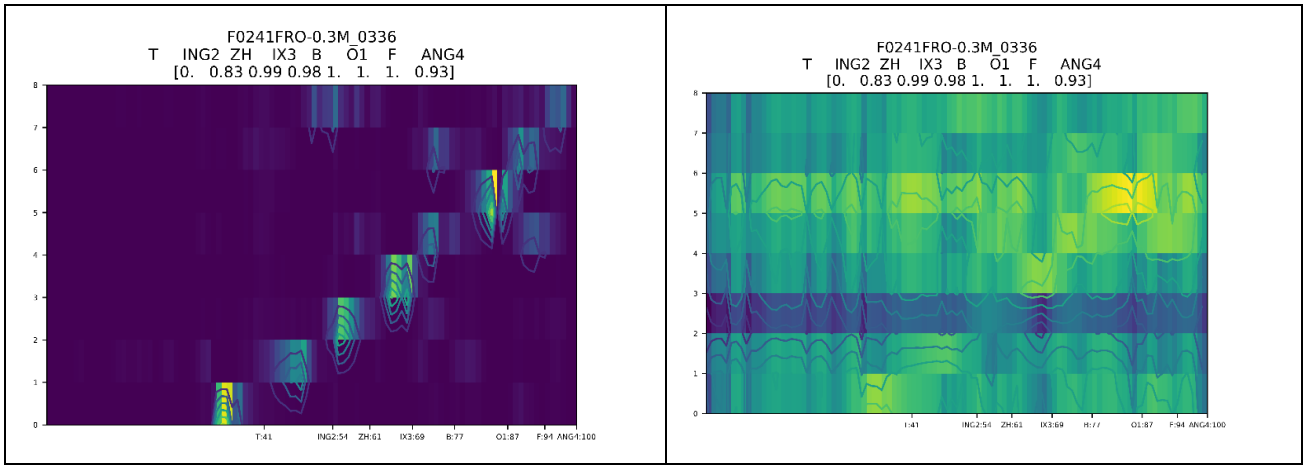
The right column is the attention output before softmax function.

FA attention weight:



Recall attention weight




F0241FRO-0.3M_0336
T ING2 ZH IX3 B Ō1 F ANG4
[0. 0.83 0.99 0.98 1. 1. 1. 0.93]
T:41
ING2:54
ZH:61
IX3:69
B:77
O1:87
F:94
ANG4:100
F0241FRO-0.3M_0336
T ING2 ZH IX3 B Ō1 F ANG4
[0. 0.83 0.99 0.98 1. 1. 1. 0.93]
I:41
ING2:54
ZH:61
IX3:69
H:77
O1:87
F:94
ANG4:100