

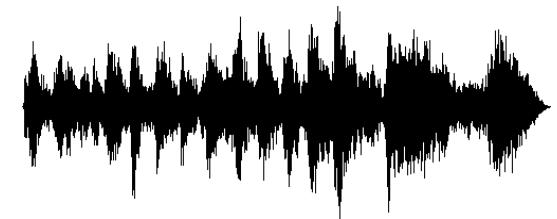
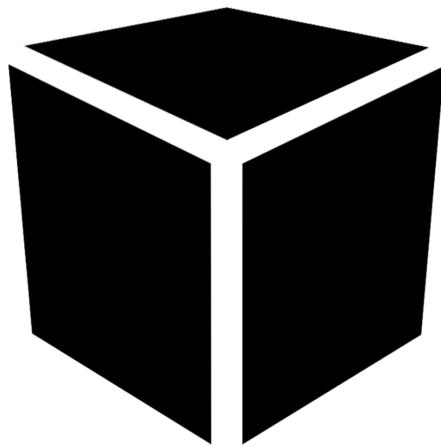
Speech Synthesis Workshop
@ Higher School of Economics
Moscow

Josh Meyer

@joshmeyerphd

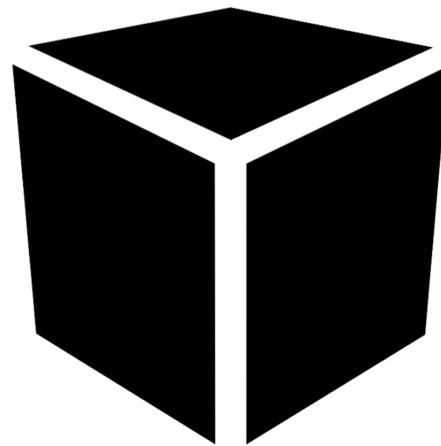
jrmeyer.github.io

Moscow always has
traffic jams.



RAW TEXT

Moscow always has
traffic jams.



RAW AUDIO

RAW TEXT

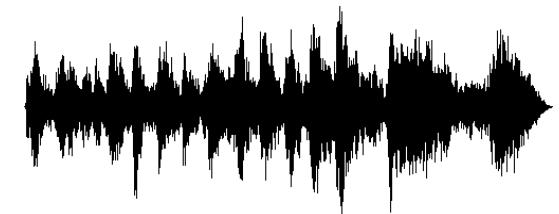
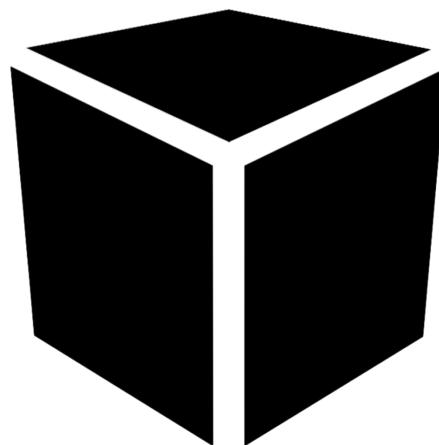


SPEECH
SYNTHESIZER

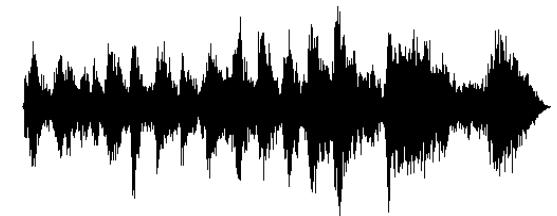
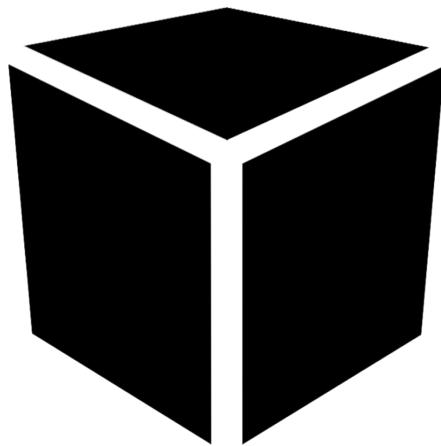


RAW AUDIO

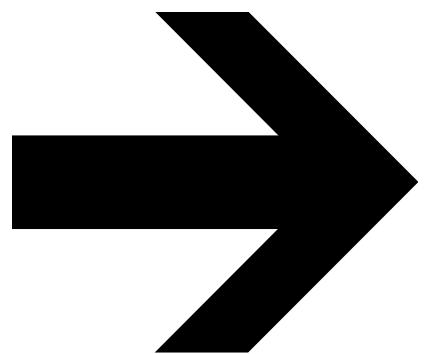
Moscow always has
traffic jams.



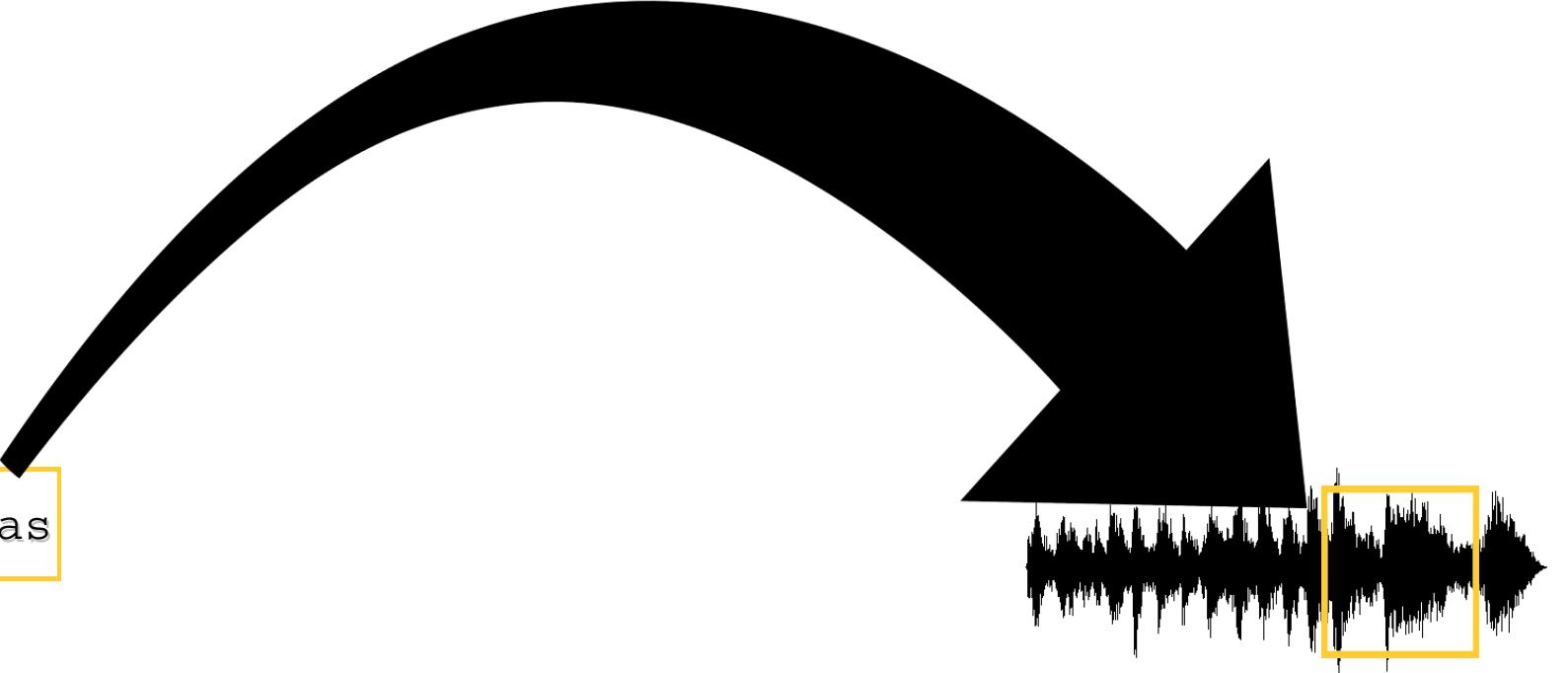
Moscow always has
traffic jams.



Moscow always has
traffic jams.



HARD

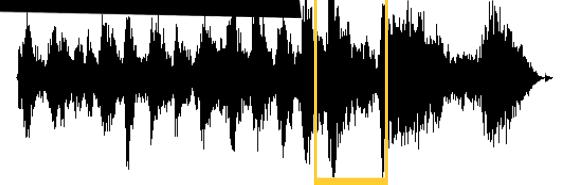


Moscow always has
traffic jams.

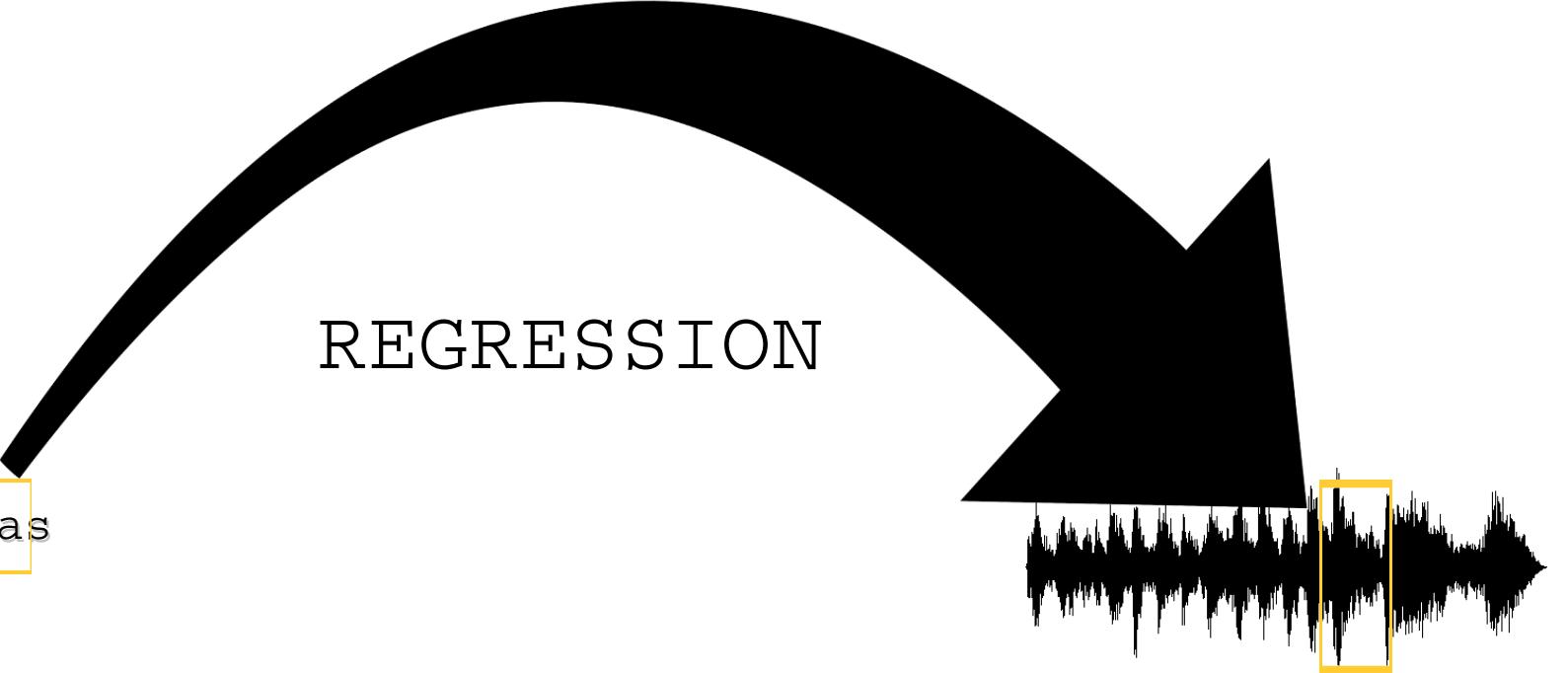
HARD



Moscow always has
traffic jams.



EASIER

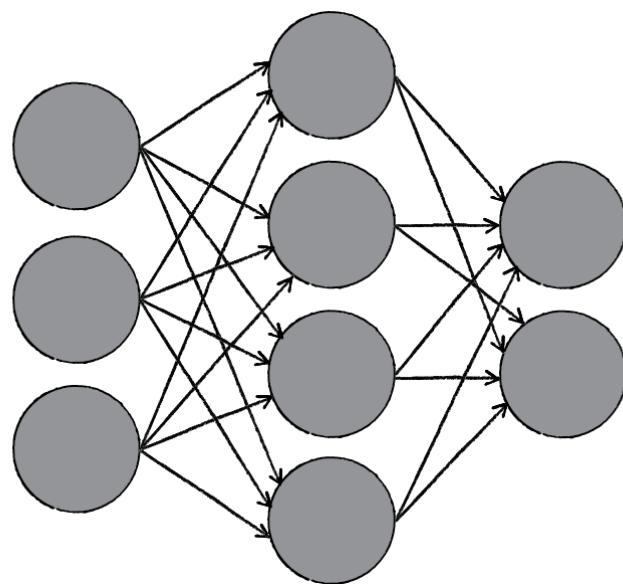


REGRESSION

Moscow always has
traffic jams.

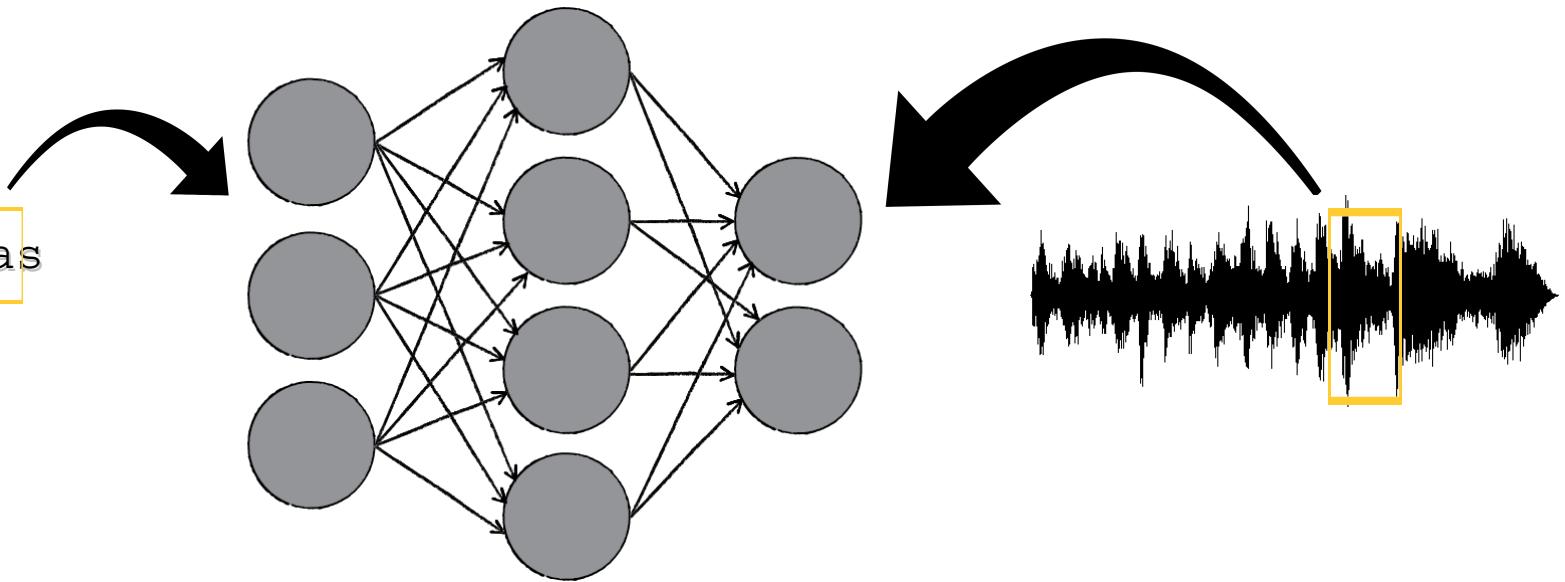


Moscow always has
traffic jams.



BECAUSE
DNNs

Moscow always has
traffic jams.

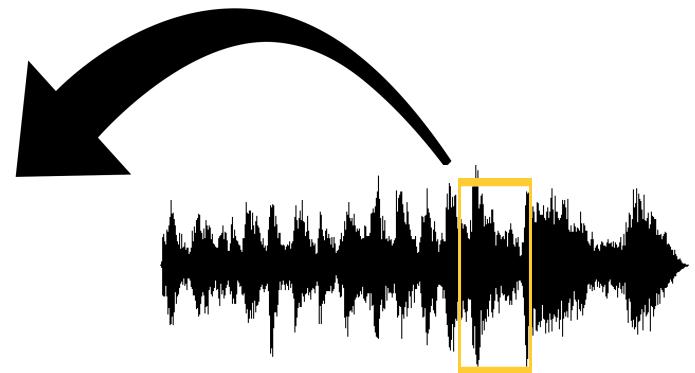


Moscow always has
traffic jams.

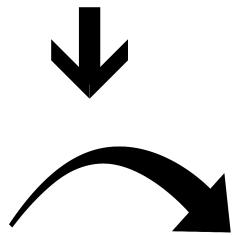


[0010000007000230100000000900]

[6010000000000008]



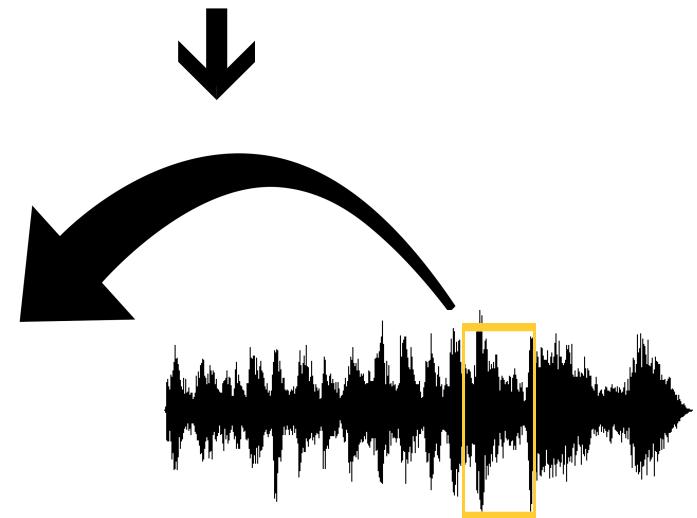
FRONTEND



Moscow always has
traffic jams.

[0010000007000230100000000900]

BACKEND



[60100000000008]

THE FRONTEND

(text feature extraction)

THE FRONTEND

1) MOTIVATION

Moscow always has
traffic jams.

Moscow always has traffic jams .



WORDS

Moscow always has
traffic jams.

M o s c o w a l w a y s h a s t r a f f i c j a m s .



LETTERS

Moscow always has
traffic jams.

M ah s k o ah L w ei z h ae z chr ae f I k dj ae m z



PHONEMES

Moscow always has
traffic jams.

[M ah s k o] [ah L w ei z] [h ae z] [chr ae f I k]
[dj ae m z]



PHONEMES + WORDS

Moscow always has
traffic jams.

[NOUN M ah s k o] [ADVERB ah L w ei z] [VERB h ae z]
[NOUN chr ae f I k] [NOUN dj ae m z]



PHONEMES + WORDS + POS

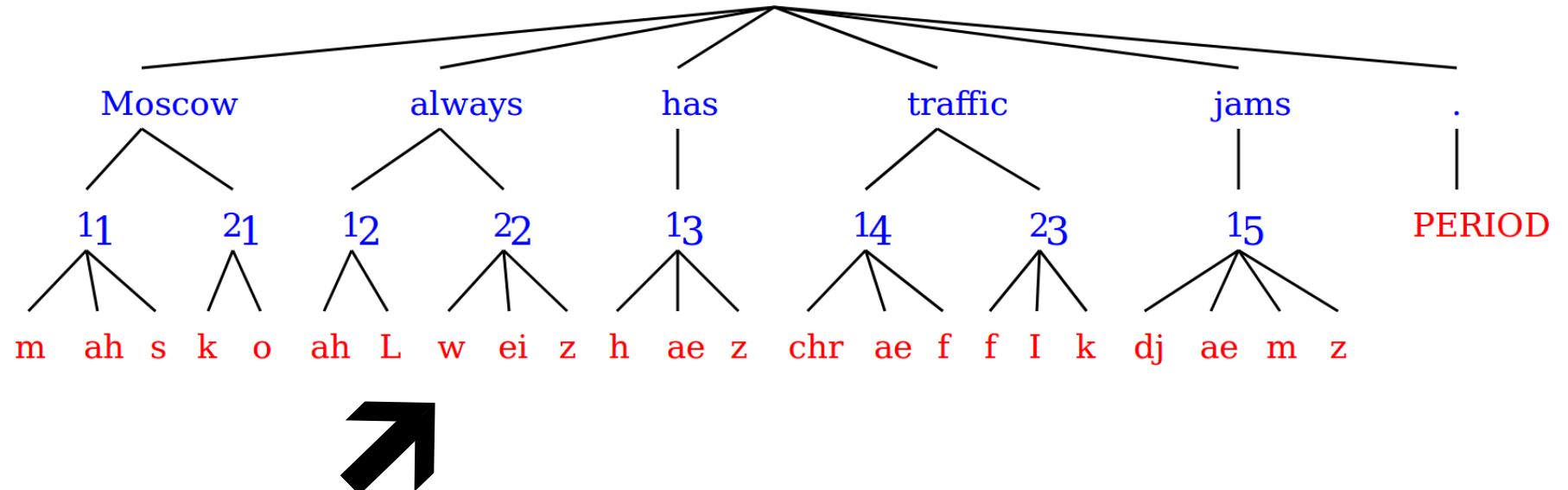
Moscow always has
traffic jams.

[NOUN [M ah] [s k o]] [ADVERB [ah L] [w ei z]]
[VERB [h ae z]] [NOUN [chr ae] [f I k]]
[NOUN [dj ae m z]]

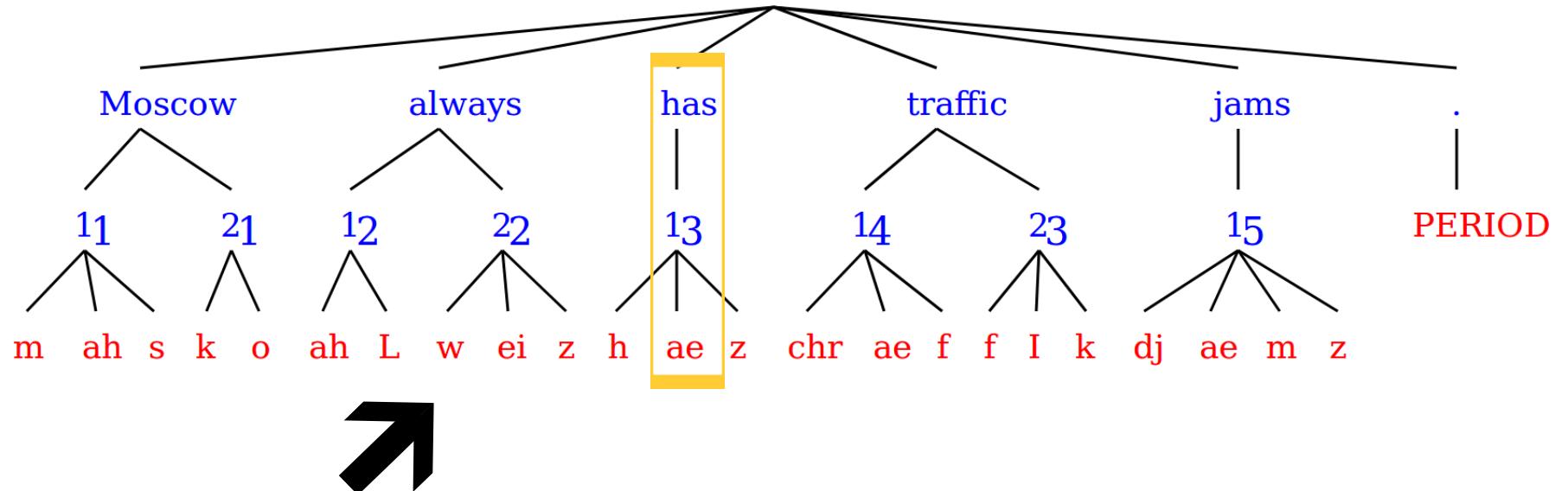


PHONEMES + WORDS + POS
+ SYLLABLES

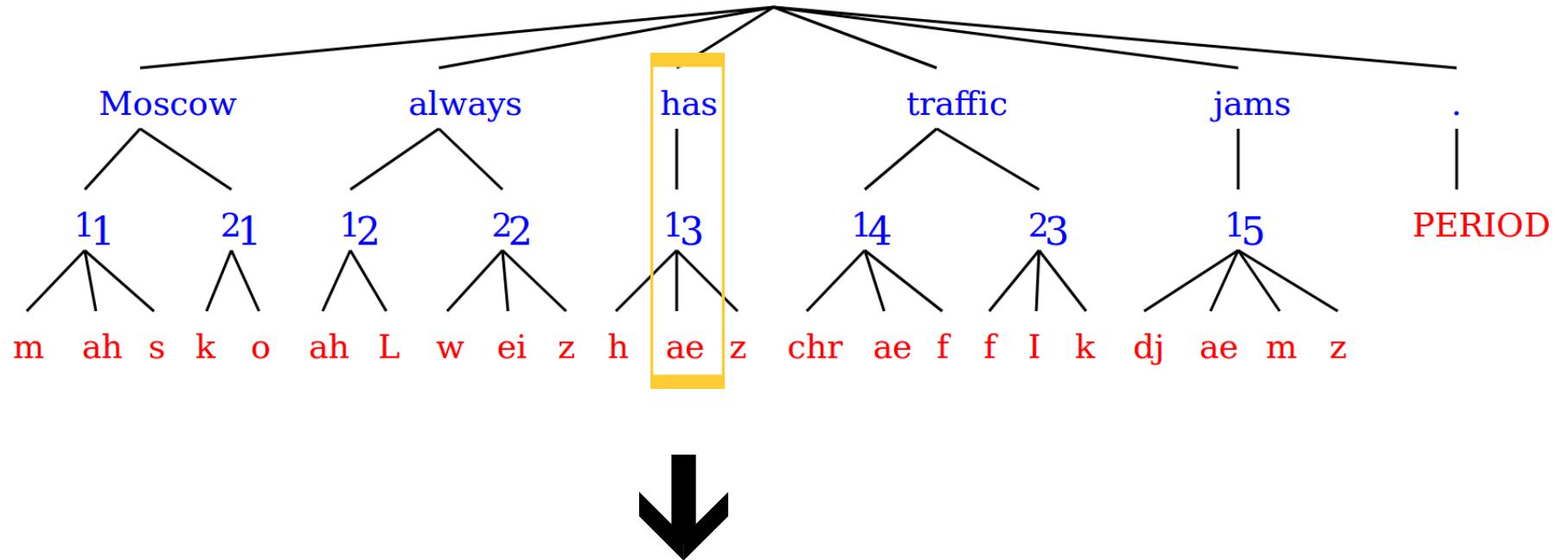
Moscow always has
traffic jams.



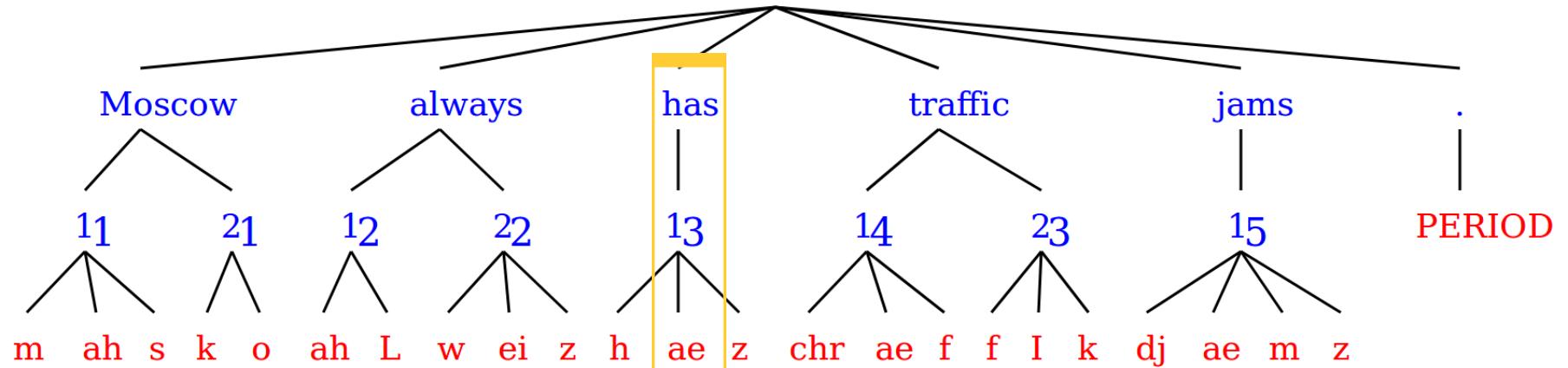
Moscow always has
traffic jams.



Moscow always has
traffic jams.



i+/0:_NA_/_1:sil/_2:ah/_3:f/_4:y/_5:0/6:8/7:4/8:1/9:8/10:0/11:9/12:0/13:0/14:76/15:0/16:
 1/17:76/18:1/19:9/20:2/21:1/22:1/23:78/24:9/25:1/26:79/27:9/28:1/29:0.6301018
 76438/30:0.685195227915/31:0.895005526298/32:0.0169384207386/33:0.0696
 863228001/34:0.184964316317/35:0.0535076519465/36:0.0302087596484/37:-
 0.148247666767/38:0.0110183463862/39:0.00683717786986/40:0.1039553511
 66/41:0.0182739382836/[6]

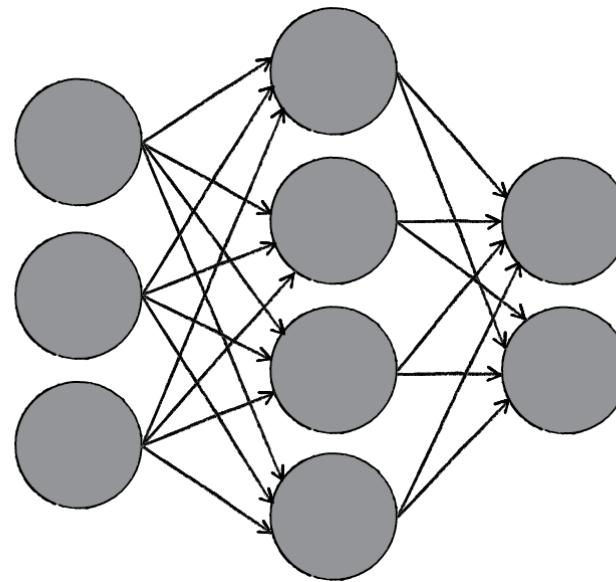
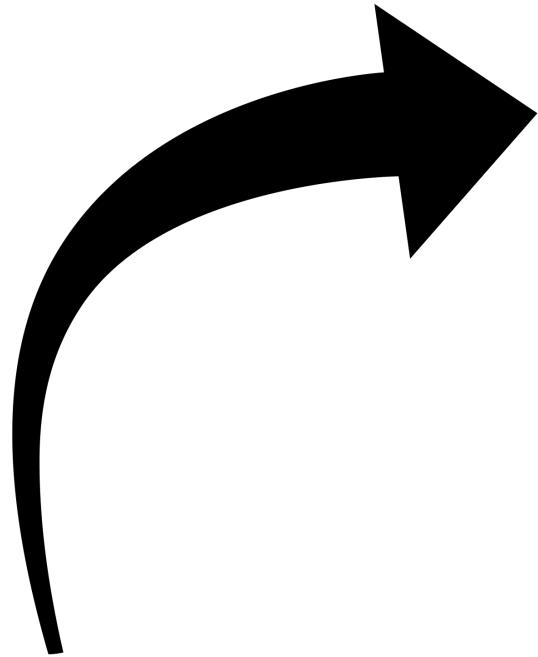


i+/0:_NA_/1:sil/2:ah/3:f/4:y/5:0/6:8/7:4/8:1/9:8/10:0/11:9/12:0/13:0/14:76/15:0/16:
1/17:76/18:1/19:9/20:2/21:1/22:1/23:78/24:9/25:1/26:79/27:9/28:1/29:0.6301018
76438/30:0.685195227915/31:0.895005526298/32:0.0169384207386/33:0.0696
863228001/34:0.184964316317/35:0.0535076519465/36:0.0302087596484/37:-
0.148247666767/38:0.0110183463862/39:0.00683717786986/40:0.1039553511
66/41:0.0182739382836/[6]

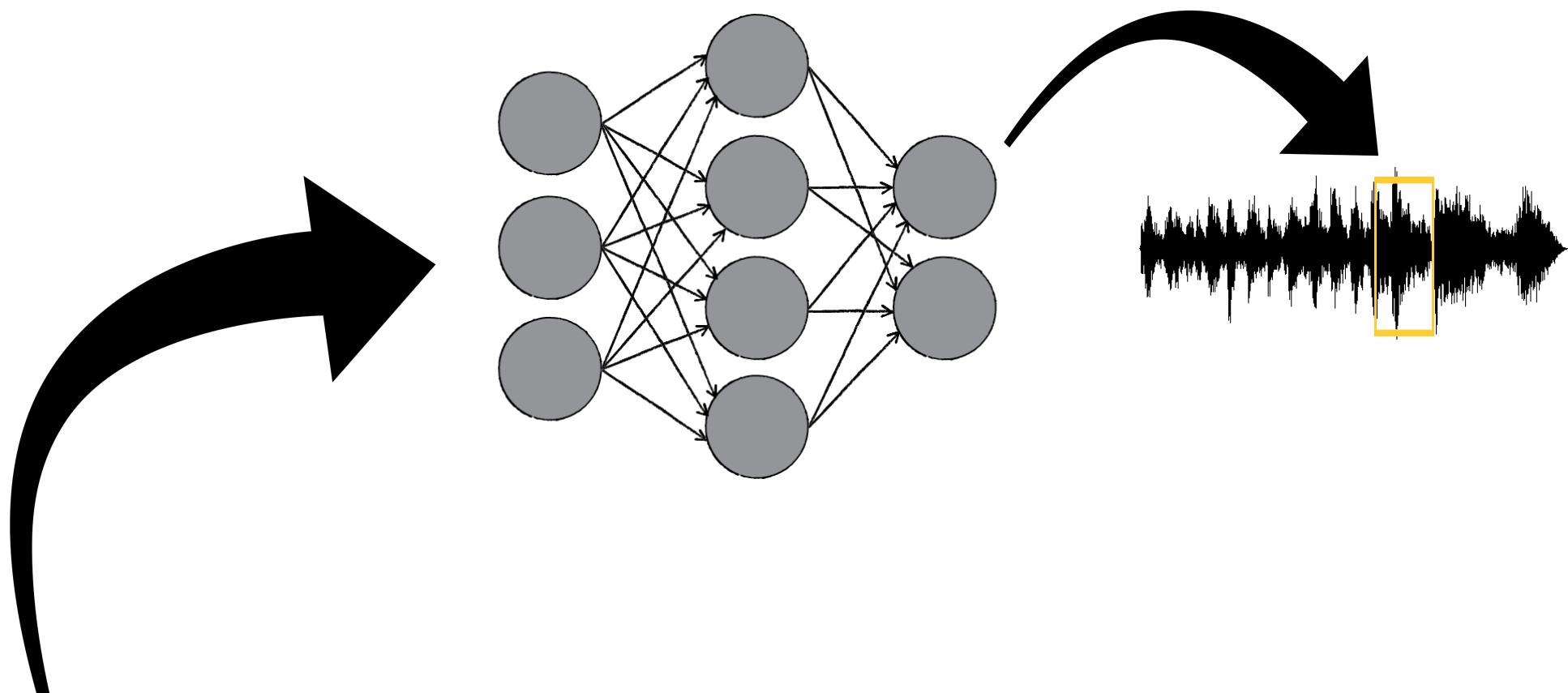


[00000100000000000001000000000000001000000000000]

[0000010000000000010000000000000100000000000]



[0000010000000000010000000000000100000000000]



```
[00000100000000000010000000000000001000000000000]
```

THE FRONTEND

1) MOTIVATION

2) TEXT → XML

Фильмсем 13 е 13:30 сөхетре пүсланең.

Фильмсем 13 е 13:30 сехетре пүсланёç.

```
<utt text="Фильмсем 13 е 13:30 сехетре пүсланёç."  
waveform="/home/ubuntu/Ossian/corpus/chv/speakers/news/wa  
v/17448-0006.wav" utterance_name="17448-0006"  
processors_used=",word_splitter,segment_adder,word_vector_t  
agger,feature_dumper,acoustic_feature_extractor,aligner">
```

```
    <token text="_END_" token_class="_END_> </token>  
    <token text="Фильмсем" token_class="word"> </token>  
    <token text=" " token_class="space"> </token>  
    <token text="13" token_class="word"> </token>  
    <token text=" " token_class="space"> </token>  
    <token text="e" token_class="word"> </token>  
    <token text=" " token_class="space"> </token>  
    <token text="13:30" token_class="word"> </token>  
    <token text=" " token_class="space"> </token>  
    <token text="сехетре" token_class="word"> </token>  
    <token text=" " token_class="space"> </token>  
    <token text="пүсланёç" token_class="word"> </token>  
    <token text"." token_class="punctuation"> </token>  
    <token text="_END_" token_class="_END_> </token>  
  
</utt>
```

Фильмсем 13 е 13:30 сехетре пүсланёç.

UTTERANCE → <utt text="Фильмсем 13 е 13:30 сехетре пүсланёç."
waveform="/home/ubuntu/Ossian/corpus/chv/speakers/news/wa
v/17448-0006.wav" utterance_name="17448-0006"
processors_used=",word_splitter,segment_adder,word_vector_t
agger,feature_dumper,acoustic_feature_extractor,aligner">

```
<token text="_END_" token_class="_END_> </token>
<token text="Фильмсем" token_class="word"> </token>
<token text=" " token_class="space"> </token>
<token text="13" token_class="word"> </token>
<token text=" " token_class="space"> </token>
<token text="e" token_class="word"> </token>
<token text=" " token_class="space"> </token>
<token text="13:30" token_class="word"> </token>
<token text=" " token_class="space"> </token>
<token text="сехетре" token_class="word"> </token>
<token text=" " token_class="space"> </token>
<token text="пүсланёç" token_class="word"> </token>
<token text"." token_class="punctuation"> </token>
<token text="_END_" token_class="_END_> </token>

</utt>
```

Фильмсем 13 е 13:30 сехетре пүсланёç.

UTTERANCE → <utt text="Фильмсем 13 е 13:30 сехетре пүсланёç."
waveform="/home/ubuntu/Ossian/corpus/chv/speakers/news/wa
v/17448-0006.wav" utterance_name="17448-0006"
processors_used=",word_splitter,segment_adder,word_vector_t
agger,feature_dumper,acoustic_feature_extractor,aligner">

<token text="_END_" token_class="_END_> </token>
<token text="Фильмсем" token_class="word"> </token>
<token text=" " token_class="space"> </token>
<token text="13" token_class="word"> </token>
<token text=" " token_class="space"> </token>
<token text="e" token_class="word"> </token>
<token text=" " token_class="space"> </token>
<token text="13:30" token_class="word"> </token>
<token text=" " token_class="space"> </token>
<token text="сехетре" token_class="word"> </token>
<token text=" " token_class="space"> </token>
<token text="пүсланёç" token_class="word"> </token>
<token text"." token_class="punctuation"> </token>
<token text="_END_" token_class="_END_> </token>

</utt>

Фильмсем 13 е 13:30 сехетре пүсланёç.

UTTERANCE → <utt text="Фильмсем 13 е 13:30 сехетре пүсланёç."
waveform="/home/ubuntu/Ossian/corpus/chv/speakers/news/wav/17448-0006.wav" utterance_name="17448-0006"
processors_used=",word_splitter,segment_adder,word_vector_tagger,feature_dumper,acoustic_feature_extractor,aligner">

PROCESS →

TOKENS →

```
<token text="_END_" token_class="_END_> </token>
<token text="Фильмсем" token_class="word"> </token>
<token text=" " token_class="space"> </token>
<token text="13" token_class="word"> </token>
<token text=" " token_class="space"> </token>
<token text="е" token_class="word"> </token>
<token text=" " token_class="space"> </token>
<token text="13:30" token_class="word"> </token>
<token text=" " token_class="space"> </token>
<token text="сехетре" token_class="word"> </token>
<token text=" " token_class="space"> </token>
<token text="пүсланёç" token_class="word"> </token>
<token text"." token_class="punctuation"> </token>
<token text="_END_" token_class="_END_> </token>

</utt>
```

Фильмсем 13 е 13:30 сехетре пүсланёç.

```
<utt text="Фильмсем 13 е 13:30 сехетре пүсланёç."  
waveform="/home/ubuntu/Ossian/corpus/chv/speakers/news/wa  
v/17448-0006.wav" utterance_name="17448-0006"  
processors_used=",word_splitter,segment_adder,word_vector_t  
agger,feature_dumper,acoustic_feature_extractor,aligner">  
  
<token text="_END_" token_class="_END_> </token>  
<token text="Фильмсем" token_class="word"> </token>  
<token text=" " token_class="space"> </token>  
<token text="13" token_class="word"> </token>  
<token text=" " token_class="space"> </token>  
<token text="e" token_class="word"> </token>  
<token text=" " token_class="space"> </token>  
<token text="13:30" token_class="word"> </token>  
<token text=" " token_class="space"> </token>  
<token text="сехетре" token_class="word"> </token>  
<token text=" " token_class="space"> </token>  
<token text="пүсланёç" token_class="word"> </token>  
<token text"." token_class="punctuation"> </token>  
<token text="_END_" token_class="_END_> </token>  
  
</utt>
```

Фильмсем 13 е 13:30 сөхетре пүсланёс.

```
<token text="Фильмсем" token_class="word"
```

```
safetext="_CYRILLICCAPITALLETTEREF__CYRILLICSMALLLETTERI__  
CYRILLICSMALLLETTEREL__CYRILLICSMALLLETTERSOTFSIGN__C  
YRILLICSMALLLETTEREM__CYRILLICSMALLLETTERES__CYRILLICS  
MALLLETTERIE__CYRILLICSMALLLETTEREM__"
```

```
vsm_d1="0.685195227915" vsm_d2="0.0696863228001"  
vsm_d3="0.0302087596484" vsm_d4="0.00683717786986"  
vsm_d5="0.0432880400823" vsm_d6="0.0132148537813" vsm_d7="-  
0.0021566008158" vsm_d8="0.000630974819014"  
vsm_d9="0.00734826504532" vsm_d10="0.00938971254099"  
>
```

```
<segment pronunciation="_CYRILLICSMALLLETTEREF_"/>  
<segment pronunciation="_CYRILLICSMALLLETTERI_"/>  
<segment pronunciation="_CYRILLICSMALLLETTEREL_"/>  
<segment pronunciation="_CYRILLICSMALLLETTERSOTFSIGN_"/>  
<segment pronunciation="_CYRILLICSMALLLETTEREM_"/>  
<segment pronunciation="_CYRILLICSMALLLETTERES_"/>  
<segment pronunciation="_CYRILLICSMALLLETTERIE_"/>  
<segment pronunciation="_CYRILLICSMALLLETTEREM_"/>
```

```
</token>
```

Фильмсем 13 е 13:30 сөхетре пүсланең.

```
<token text="Фильмсем" token_class="word"
```

```
safetext=_CYRILLICCAPITALLETTEREF__CYRILLICSMALLLETTERI__  
SAFETXT → CYRILLICSMALLLETTEREL__CYRILLICSMALLLETTERSOTSIGN__C  
YRILLICSMALLLETTEREM__CYRILLICSMALLLETTERES__CYRILLICS  
MALLLETTERIE__CYRILLICSMALLLETTEREM__
```

```
vsm_d1="0.685195227915" vsm_d2="0.0696863228001"  
vsm_d3="0.0302087596484" vsm_d4="0.00683717786986"  
vsm_d5="0.0432880400823" vsm_d6="0.0132148537813" vsm_d7="-  
0.0021566008158" vsm_d8="0.000630974819014"  
vsm_d9="0.00734826504532" vsm_d10="0.00938971254099"  
>
```

```
<segment pronunciation="_CYRILLICSMALLLETTEREF_"/>  
<segment pronunciation="_CYRILLICSMALLLETTERI_"/>  
<segment pronunciation="_CYRILLICSMALLLETTEREL_"/>  
<segment pronunciation="_CYRILLICSMALLLETTERSOTSIGN_"/>  
<segment pronunciation="_CYRILLICSMALLLETTEREM_"/>  
<segment pronunciation="_CYRILLICSMALLLETTERES_"/>  
<segment pronunciation="_CYRILLICSMALLLETTERIE_"/>  
<segment pronunciation="_CYRILLICSMALLLETTEREM_"/>
```

```
</token>
```

Фильмсем 13 е 13:30 сөхетре пүсланең.

<token text="Фильмсем" token_class="word"

safetext=_CYRILLICCAPITALLETTEREF__CYRILLICSMALLLETTERI__
CYRILLICSMALLLETTEREL__CYRILLICSMALLLETTERSOTFSIGN__C
YRILLICSMALLLETTEREM__CYRILLICSMALLLETTERES__CYRILLICS
MALLLETTERIE__CYRILLICSMALLLETTEREM__

vsm_d1="0.685195227915" vsm_d2="0.0696863228001"
vsm_d3="0.0302087596484" vsm_d4="0.00683717786986"

VECTORS → vsm_d5="0.0432880400823" vsm_d6="0.0132148537813" vsm_d7="-
0.0021566008158" vsm_d8="0.000630974819014"
vsm_d9="0.00734826504532" vsm_d10="0.00938971254099"
>

<segment pronunciation="_CYRILLICSMALLLETTEREF_"/>
<segment pronunciation="_CYRILLICSMALLLETTERI_"/>
<segment pronunciation="_CYRILLICSMALLLETTEREL_"/>
<segment pronunciation="_CYRILLICSMALLLETTERSOTFSIGN_"/>
<segment pronunciation="_CYRILLICSMALLLETTEREM_"/>
<segment pronunciation="_CYRILLICSMALLLETTERES_"/>
<segment pronunciation="_CYRILLICSMALLLETTERIE_"/>
<segment pronunciation="_CYRILLICSMALLLETTEREM_"/>

</token>

Фильмсем 13 е 13:30 сөхетре пүсланең.

<token text="Фильмсем" token_class="word"

safetext=_CYRILLICCAPITALLETTEREF__CYRILLICSMALLLETTERI__
CYRILLICSMALLLETTEREL__CYRILLICSMALLLETTERSOTFSIGN__C
YRILLICSMALLLETTEREM__CYRILLICSMALLLETTERES__CYRILLICS
MALLLETTERIE__CYRILLICSMALLLETTEREM__

vsm_d1="0.685195227915" vsm_d2="0.0696863228001"
vsm_d3="0.0302087596484" vsm_d4="0.00683717786986"

VECTORS → vsm_d5="0.0432880400823" vsm_d6="0.0132148537813" vsm_d7="-
0.0021566008158" vsm_d8="0.000630974819014"
vsm_d9="0.00734826504532" vsm_d10="0.00938971254099"
>

PHONES → <segment pronunciation="_CYRILLICSMALLLETTEREF_"/>
<segment pronunciation="_CYRILLICSMALLLETTERI_"/>
<segment pronunciation="_CYRILLICSMALLLETTEREL_"/>
<segment pronunciation="_CYRILLICSMALLLETTERSOTFSIGN_"/>
<segment pronunciation="_CYRILLICSMALLLETTEREM_"/>
<segment pronunciation="_CYRILLICSMALLLETTERES_"/>
<segment pronunciation="_CYRILLICSMALLLETTERIE_"/>
<segment pronunciation="_CYRILLICSMALLLETTEREM_"/>

</token>

Фильмсем 13 е 13:30 сөхетре пүсланең.

<token text="13" token_class="word"

SAFETXT → safetext="_DIGITONE__DIGITTHREE_"

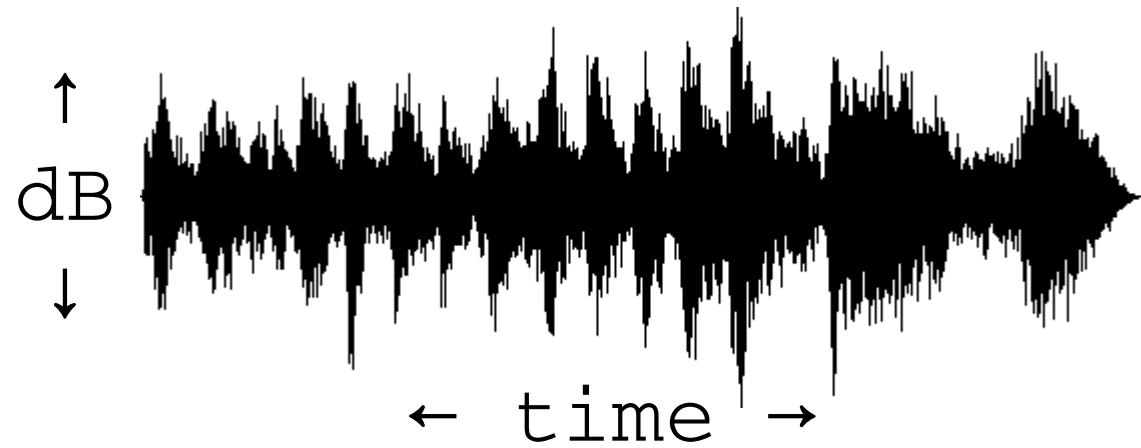
vsm_d1="0.175666404262" vsm_d2="0.152223466799"
vsm_d3="-0.0307015100812" vsm_d4="0.0999614943211"
vsm_d5="0.0698700498727" vsm_d6="0.25982222856"
vsm_d7="0.0213851732067" vsm_d8="-0.00852679385956"
vsm_d9="-0.0024125017034" vsm_d10="0.0225874466051"
>

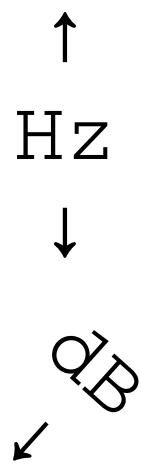
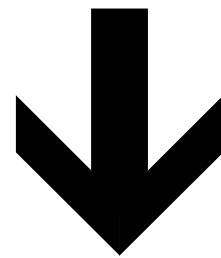
VECTORS → <segment pronunciation="_DIGITONE_"/>
<segment pronunciation="_DIGITTHREE_"/>
</token>

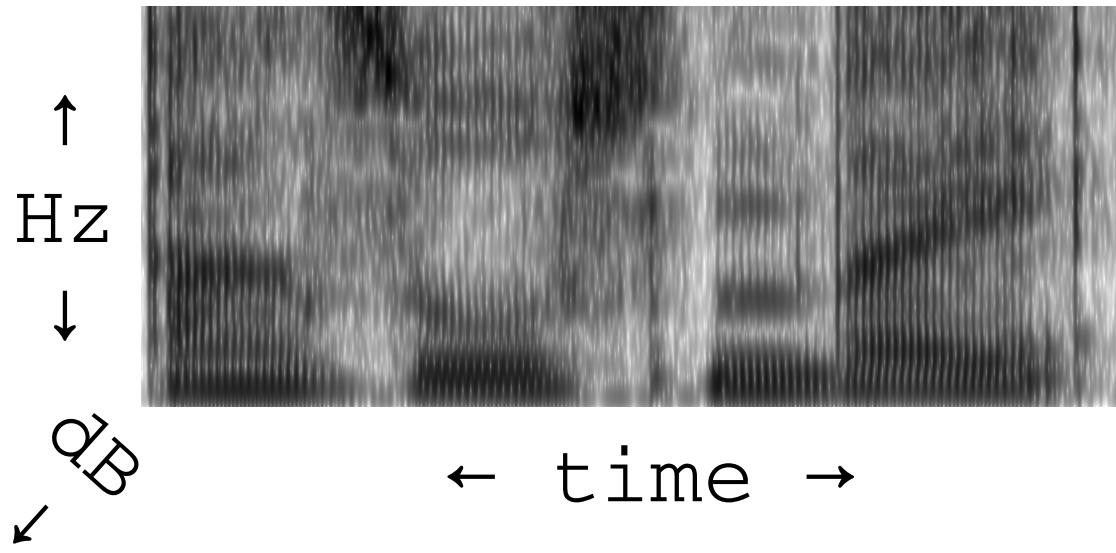
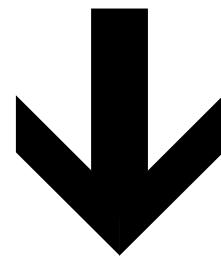
THE BACKEND

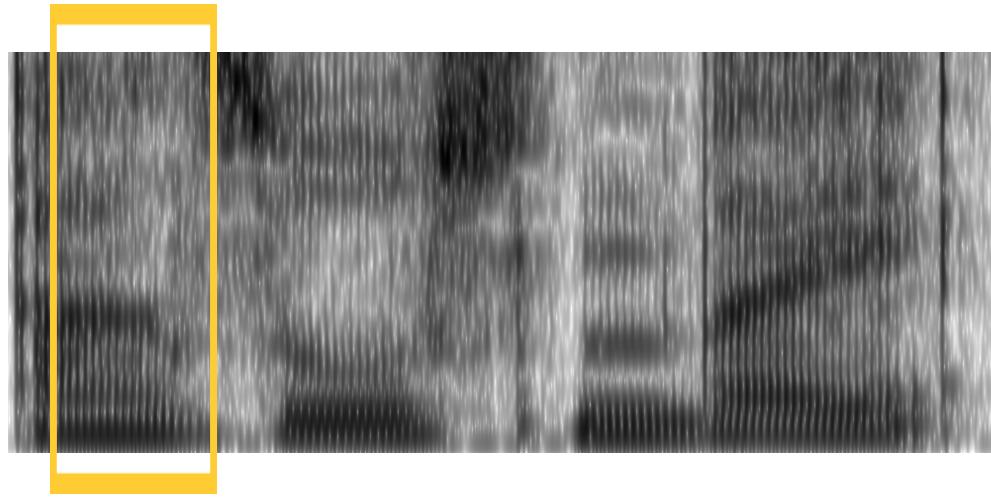
(audio feature extraction)

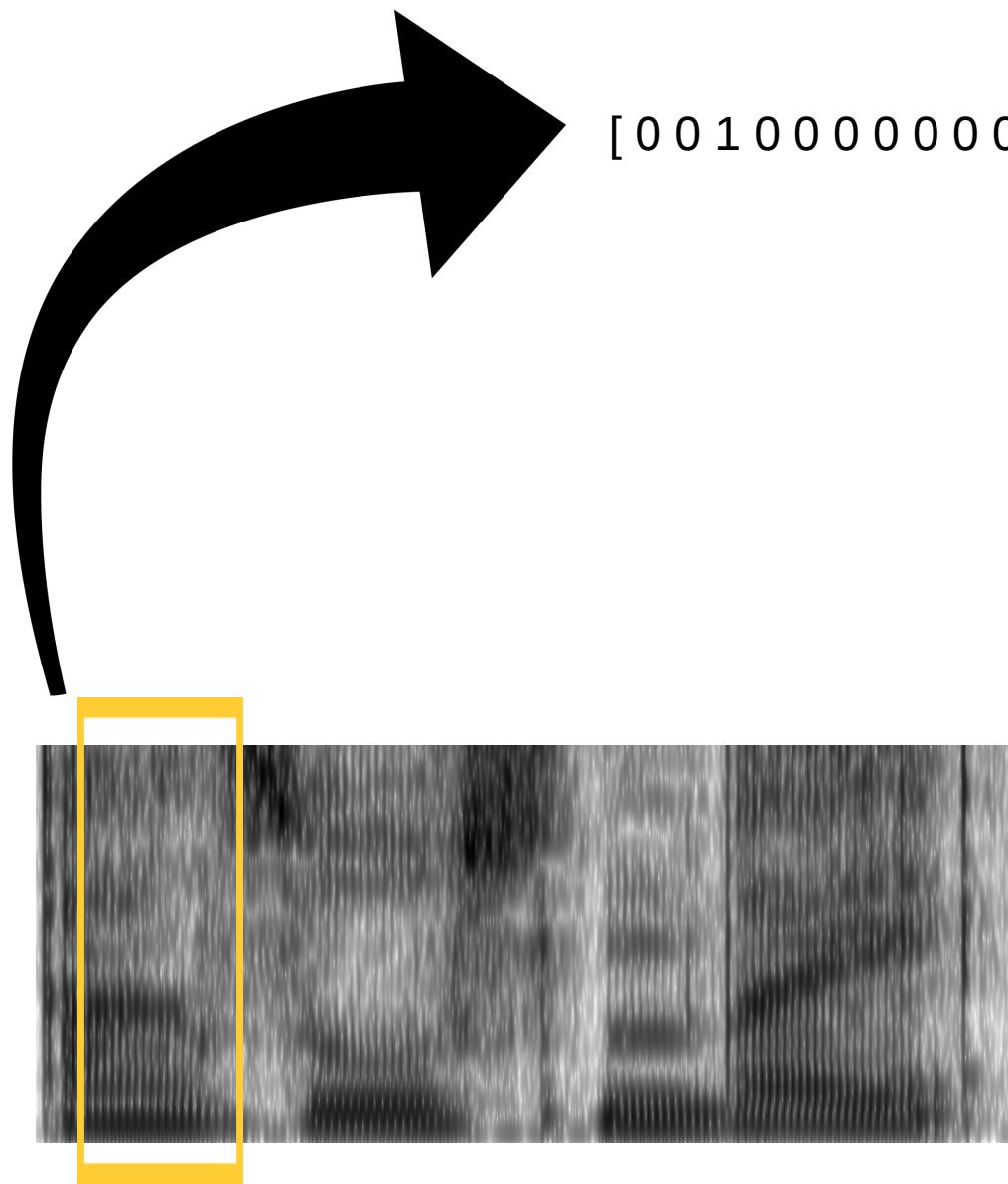












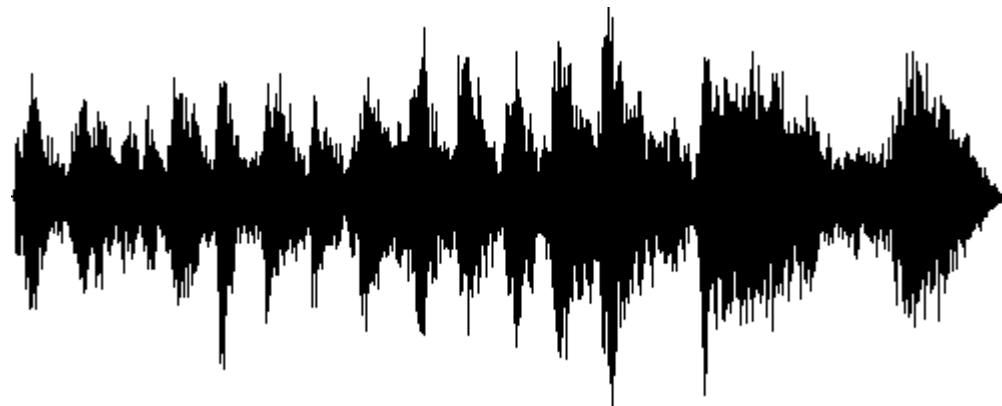
[0010000000000010000000]

REGRESSION

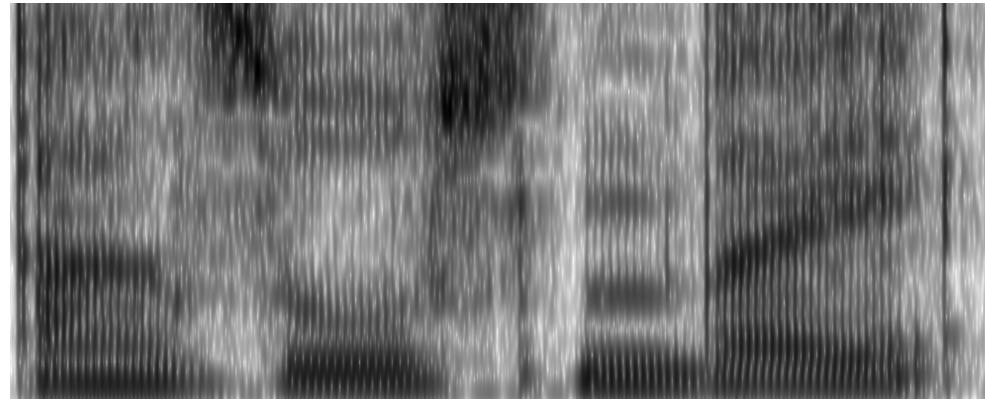
REGRESSION

1) LABELED DATA

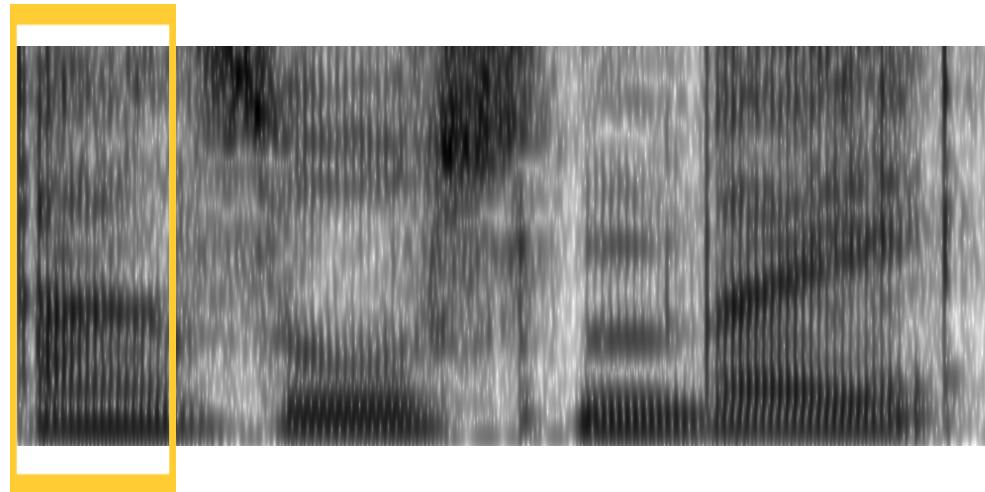
Moscow always has traffic jams.



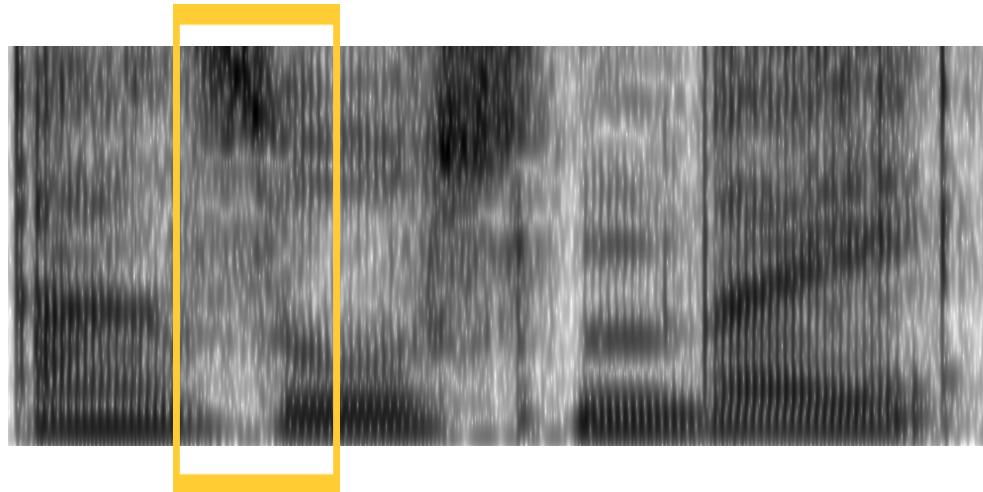
Moscow always has traffic jams.



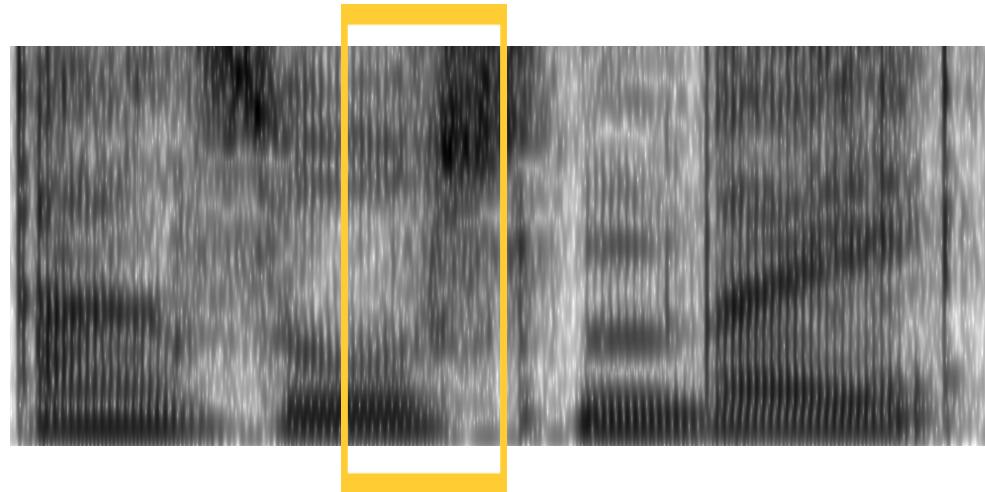
Moscow always has traffic jams.



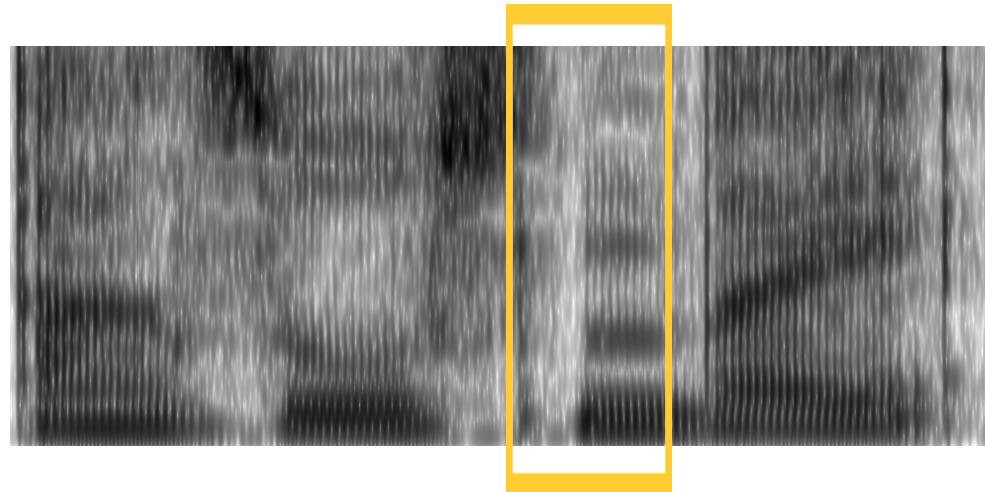
Moscow always has traffic jams.



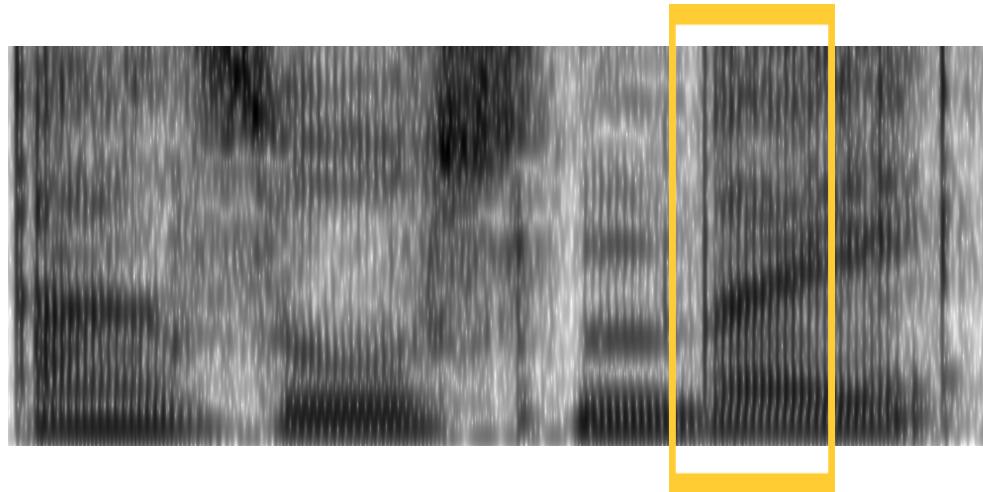
Moscow always has traffic jams.



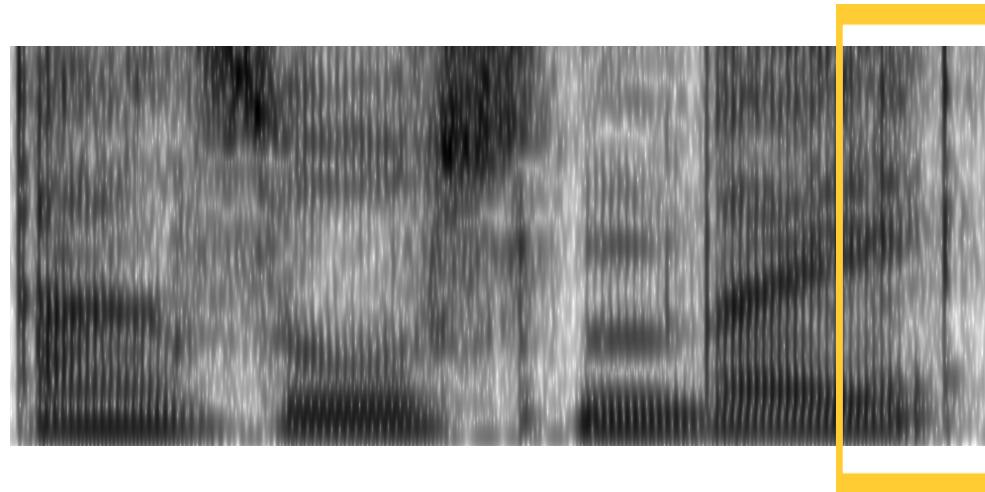
Moscow always has traffic jams.



Moscow always has traffic jams.

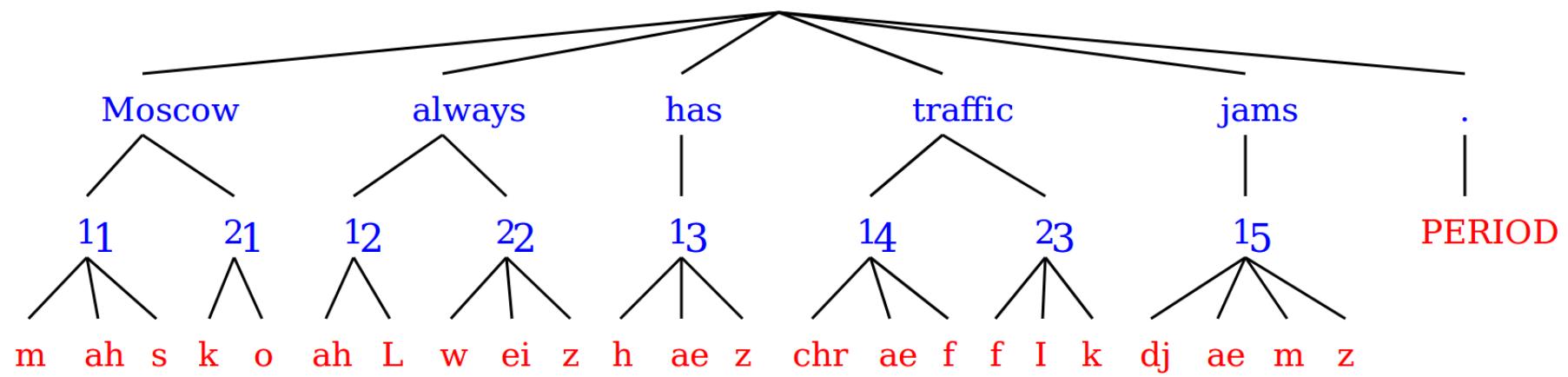


Moscow always has traffic jams.



Moscow always has traffic jams.

```
[0 0 1 0 0 0 0 1 0 0 0 0 0 0 0  
[0 0 2 0 0 0 3 0 0 0 7 0 0 0 2  
[3 0 1 0 0 0 0 0 0 0 9 0 0 0 0  
[0 0 5 0 9 0 0 1 0 0 0 4 0 0 0  
[6 0 1 0 0 0 0 1 0 0 0 0 0 0 8  
[0 0 2 0 0 0 3 0 0 9 0 0 0 0 0  
[0 0 1 0 0 0 0 0 0 0 1 0 0 0 0  
[0 3 0 0 0 0 0 1 0 0 7 0 0 0 2  
[0 0 5 0 9 0 4 0 0 0 0 4 0 0 0  
[3 0 1 0 0 0 9 0 0 0 0 0 0 0 0  
[6 0 1 0 0 0 0 0 0 0 0 0 0 0 0  
[0 0 2 0 0 0 3 0 0 9 0 0 0 0 0  
[0 0 0 0 0 0 1 0 0 0 0 0 0 0 0  
[6 0 2 0 0 0 1 0 0 7 0 0 0 2 0  
[0 0 0 0 0 0 1 0 0 0 0 0 0 0 0  
[0 0 2 0 0 0 1 0 0 0 0 0 0 0 0  
[3 0 1 0 0 0 9 0 0 0 0 0 0 0 0  
[0 0 5 0 9 0 4 1 0 0 0 4 0 0 0  
[6 0 1 0 0 0 1 0 0 0 0 0 0 0 0  
[0 0 2 0 0 0 1 0 0 0 0 0 0 0 0  
[0 0 0 0 0 0 1 0 0 0 0 0 0 0 0  
[6 0 3 0 0 0 0 1 0 0 7 0 0 0 2  
[0 0 5 0 9 0 4 1 0 0 0 4 0 0 0  
[6 0 1 0 0 0 1 0 0 0 0 0 0 0 0  
[0 0 2 0 0 0 3 1 0 0 0 0 0 0 0]
```



i+/0:_NA_/_1:sil/_2:ah/_3:f/_4:y/_5:0/_6:8/_7:4/_8:1/_9:8/_10:0/_11:9/_12:0/_13:0/_14:76/_15:0/_16:1/_17:76/_18:1/_19:9/_20:2/_21:1/_22:1/_23:78/_24:9/_25:1/_26:79/_27:9/_28:1/_29:0.630101876438/_30:0.685195227915/_31:0.895005526298/_32:0.0169384207386/_33:0.0696863228001/_34:0.18496

i+/0:_NA_/1:sil/2:ah/3:f/4:y/5:0/6:8/7:4/8:1/9:8/10:0/11:9/12:0/13:0/14:76/15:0/16:1/17:76/18:1/19:9/20:2/21:1/22:1/23:78/24:9/25:1/26:79/27:9/28:1/29:0.630101876438/30:0.685195227915/31:0.895005526298/32:0.0169384207386/33:0.0696863228001/34:0.18496

i+/0:_NA_/1:sil/2:ah/3:f/4:y/5:0/6:8/7:4/8:1/9:8/10:0/11:9/12:0/13:0/14:76/15:0/16:1/17:76/18:1/19:9/20:2/21:1/22:1/23:78/24:9/25:1/26:79/27:9/28:1/29:0.630101876438/30:0.685195227915/31:0.895005526298/32:0.0169384207386/33:0.0696863228001/34:0.18496

[000001000000000001000080000500100000000000]

[000001004000000001000070000000100000000000]

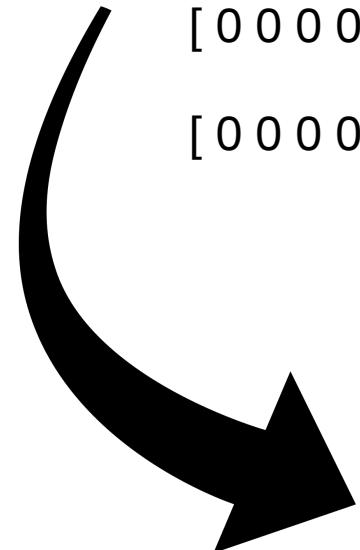
[000001000050000001000000009000100000000000]

[001000010000000
[002000300070002
[30100000009000
[005090010004000
[601000100000800
[002000300900000
[0010000001000100
[030000010070002
[005090400004000
[601000000000000
[002000300900000
[000000100000000
[002000300900000
[000000100000000
[005090400004000
[601000000100000
[002000010000000
[000000100000000
[005090410004000
[601000010000000
[002000310000000

[000001000000000001000080000500100000000000]

[00000100400000000100007000000010000000000]

[00000100005000000100000000900010000000000]

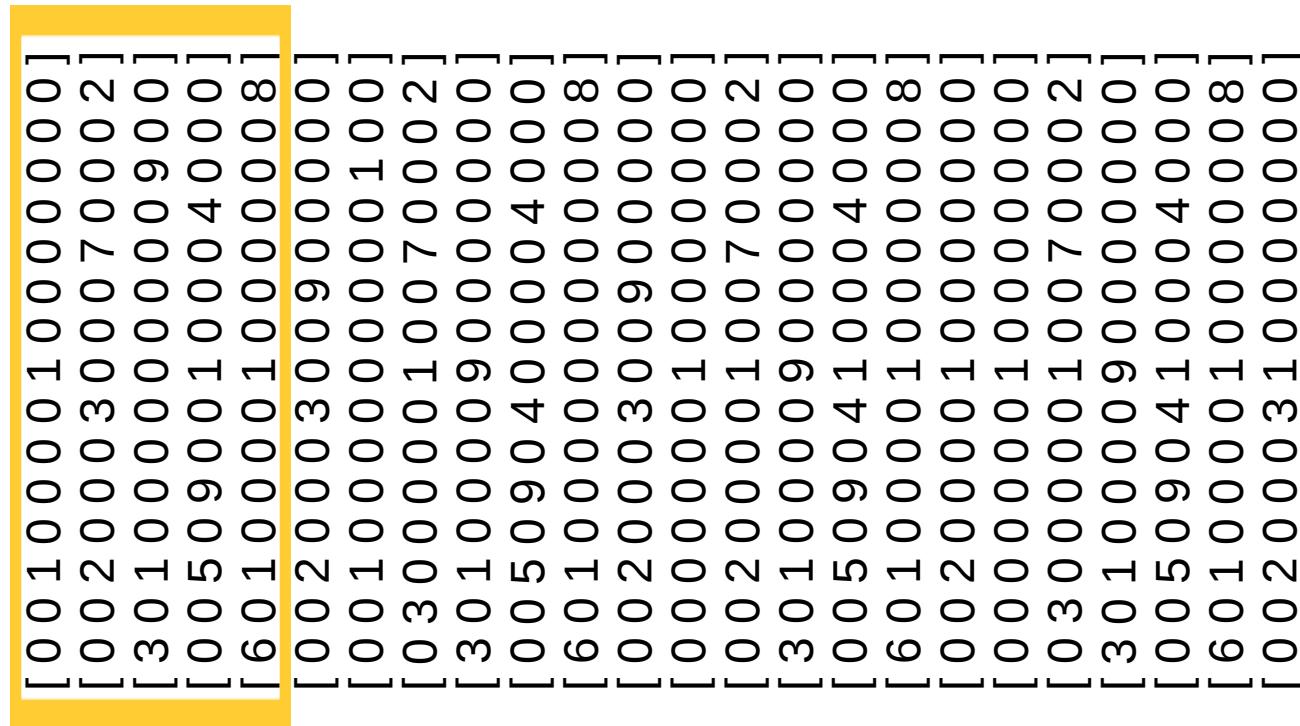


[00000100000000000100008000050010000000000]

[00000100400000000100007000000010000000000]

[000001000050000001000000009000100000000000]

HTK ALIGNMENT



WORD →

```
<token text="иlhě" token_class="word"
safetext="_CYRILLICSMALLLETTERI__CYRILLICSMALLLETTEREL__CYRILLICSMALLLETTEREN_
_CYRILLICSMALLLETTERIEWITHBREVE_" vsm_d1="0.718258053926" vsm_d2="0.84356178952"
vsm_d3="0.245865192265" vsm_d4="-0.154229921545" vsm_d5="0.0666986423832" vsm_d6="-
0.0339937700926" vsm_d7="-0.0034077854402" vsm_d8="-0.0167915876804" vsm_d9="-
0.00178861541477" vsm_d10="-0.00551910881847" start="2455" end="2765" has_silence="no"
phrase_start="False" phrase_end="True">

<segment pronunciation="_CYRILLICSMALLLETTERI_" start="2455" end="2510">
  <state start="2455" end="2460"/>
  <state start="2460" end="2475"/>
  <state start="2475" end="2500"/>
  <state start="2500" end="2505"/>
  <state start="2505" end="2510"/>
</segment>

<segment pronunciation="_CYRILLICSMALLLETTEREL_" start="2510" end="2565">
  <state start="2510" end="2515"/>
  <state start="2515" end="2520"/>
  <state start="2520" end="2525"/>
  <state start="2525" end="2550"/>
  <state start="2550" end="2565"/>
</segment>

<segment pronunciation="_CYRILLICSMALLLETTEREN_" start="2565" end="2665">
  <state start="2565" end="2645"/>
  <state start="2645" end="2650"/>
  <state start="2650" end="2655"/>
  <state start="2655" end="2660"/>
  <state start="2660" end="2665"/>
</segment>

<segment pronunciation="_CYRILLICSMALLLETTERIEWITHBREVE_" start="2665" end="2765">
  <state start="2665" end="2730"/>
  <state start="2730" end="2750"/>
  <state start="2750" end="2755"/>
  <state start="2755" end="2760"/>
  <state start="2760" end="2765"/>
</segment>
</token>
```

WORD → <token text="иљё" token_class="word"
safetext="_CYRILLICSMALLLETTERI_CYRILLICSMALLLETTEREL_CYRILLICSMALLLETTEREN_
CYRILLICSMALLLETTERIEWITHBREVE" vsm_d1="0.718258053926" vsm_d2="0.84356178952"
vsm_d3="0.245865192265" vsm_d4="-0.154229921545" vsm_d5="0.0666986423832" vsm_d6="-
0.0339937700926" vsm_d7="-0.0034077854402" vsm_d8="-0.0167915876804" vsm_d9="-
0.00178861541477" vsm_d10="-0.00551910881847" start="2455" end="2765" has_silence="no"
phrase_start="False" phrase_end="True">

PHONEME → <segment pronunciation="_CYRILLICSMALLLETTERI_" start="2455" end="2510">
 <state start="2455" end="2460"/>
 <state start="2460" end="2475"/>
 <state start="2475" end="2500"/>
 <state start="2500" end="2505"/>
 <state start="2505" end="2510"/>
</segment>

PHONEME → <segment pronunciation="_CYRILLICSMALLLETTEREL_" start="2510" end="2565">
 <state start="2510" end="2515"/>
 <state start="2515" end="2520"/>
 <state start="2520" end="2525"/>
 <state start="2525" end="2550"/>
 <state start="2550" end="2565"/>
</segment>

PHONEME → <segment pronunciation="_CYRILLICSMALLLETTEREN_" start="2565" end="2665">
 <state start="2565" end="2645"/>
 <state start="2645" end="2650"/>
 <state start="2650" end="2655"/>
 <state start="2655" end="2660"/>
 <state start="2660" end="2665"/>
</segment>

PHONEME → <segment pronunciation="_CYRILLICSMALLLETTERIEWITHBREVE_" start="2665" end="2765">
 <state start="2665" end="2730"/>
 <state start="2730" end="2750"/>
 <state start="2750" end="2755"/>
 <state start="2755" end="2760"/>
 <state start="2760" end="2765"/>
</segment>
</token>

WORD → <token text="иљё" token_class="word"
safetext="_CYRILLICSMALLLETTERI_CYRILLICSMALLLETTEREL_CYRILLICSMALLLETTEREN_
CYRILLICSMALLLETTERIEWITHBREVE" vsm_d1="0.718258053926" vsm_d2="0.84356178952"
vsm_d3="0.245865192265" vsm_d4="-0.154229921545" vsm_d5="0.0666986423832" vsm_d6="-
0.0339937700926" vsm_d7="-0.0034077854402" vsm_d8="-0.0167915876804" vsm_d9="-
0.00178861541477" vsm_d10="-0.00551910881847" start="2455" end="2765" has_silence="no"
phrase_start="False" phrase_end="True">

PHONEME → <segment pronunciation="_CYRILLICSMALLLETTERI_" start="2455" end="2510">
<state start="2455" end="2460"/>
<state start="2460" end="2475"/> ← TIME STAMPS
<state start="2475" end="2500"/>
<state start="2500" end="2505"/>
<state start="2505" end="2510"/>
</segment>

PHONEME → <segment pronunciation="_CYRILLICSMALLLETTEREL_" start="2510" end="2565">
<state start="2510" end="2515"/>
<state start="2515" end="2520"/>
<state start="2520" end="2525"/> ← TIME STAMPS
<state start="2525" end="2550"/>
<state start="2550" end="2565"/>
</segment>

PHONEME → <segment pronunciation="_CYRILLICSMALLLETTEREN_" start="2565" end="2665">
<state start="2565" end="2645"/>
<state start="2645" end="2650"/>
<state start="2650" end="2655"/> ← TIME STAMPS
<state start="2655" end="2660"/>
<state start="2660" end="2665"/>
</segment>

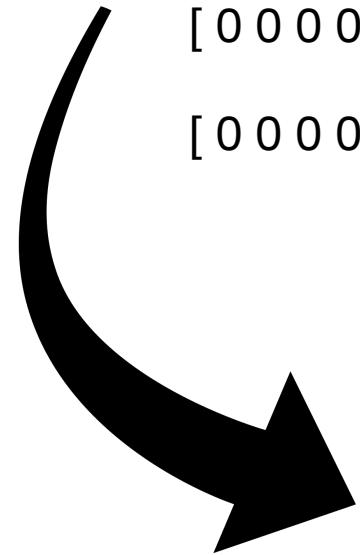
PHONEME → <segment pronunciation="_CYRILLICSMALLLETTERIEWITHBREVE_" start="2665" end="2765">
<state start="2665" end="2730"/>
<state start="2730" end="2750"/>
<state start="2750" end="2755"/> ← TIME STAMPS
<state start="2755" end="2760"/>
<state start="2760" end="2765"/>
</segment>
</token>

[000001000000000001000080000500100000000000]

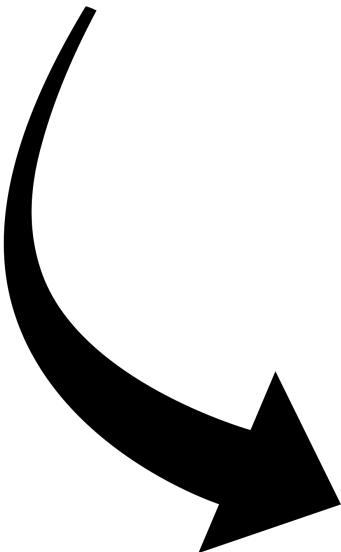
[000001004000000001000070000000100000000000]

[000001000050000001000000009000100000000000]

[0010000100000000]	[0020003009000000]	[0010000000010000]	[0300000100700020]	[3010000900000000]	[0050904000040000]	[6010000000000000]	[0020003009000000]	[0000001000000000]	[0200010000000000]	[3010000900000000]	[0050904100040000]	[6010000100000000]	[0020001000000000]	[3000001000000000]	[0050904100040000]	[6010000100000000]	[0020003100000000]
[0020003009000000]	[0010000000010000]	[0300000100700020]	[3010000900000000]	[0050904000040000]	[6010000000000000]	[0020003009000000]	[0000001000000000]	[0200010000000000]	[3010000900000000]	[0050904100040000]	[6010000100000000]	[0020001000000000]	[3000001000000000]	[0050904100040000]	[6010000100000000]	[0020003100000000]	
[0010000000010000]	[0300000100700020]	[3010000900000000]	[0050904000040000]	[6010000000000000]	[0020003009000000]	[0000001000000000]	[0200010000000000]	[3010000900000000]	[0050904100040000]	[6010000100000000]	[0020001000000000]	[3000001000000000]	[0050904100040000]	[6010000100000000]	[0020003100000000]		
[0300000100700020]	[3010000900000000]	[0050904000040000]	[6010000000000000]	[0020003009000000]	[0000001000000000]	[0200010000000000]	[3010000900000000]	[0050904100040000]	[6010000100000000]	[0020001000000000]	[3000001000000000]	[0050904100040000]	[6010000100000000]	[0020003100000000]			
[3010000900000000]	[0050904000040000]	[6010000000000000]	[0020003009000000]	[0000001000000000]	[0200010000000000]	[3010000900000000]	[0050904100040000]	[6010000100000000]	[0020001000000000]	[3000001000000000]	[0050904100040000]	[6010000100000000]	[0020003100000000]				
[0050904000040000]	[6010000000000000]	[0020003009000000]	[0000001000000000]	[0200010000000000]	[3010000900000000]	[0050904100040000]	[6010000100000000]	[0020001000000000]	[3000001000000000]	[0050904100040000]	[6010000100000000]	[0020003100000000]					
[6010000000000000]	[0020003009000000]	[0000001000000000]	[0200010000000000]	[3010000900000000]	[0050904100040000]	[6010000100000000]	[0020001000000000]	[3000001000000000]	[0050904100040000]	[6010000100000000]	[0020003100000000]						

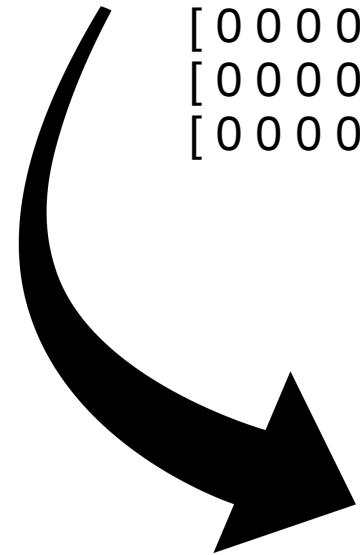


[00000100000000000100008000050010000000000]



[001000010000000
[002000300070002
[30100000009000
[005090010004000
[601000010000008]

```
[0000010000000000001000080000500100000000000]  
[0000010000000000001000080000500100000000000]  
[0000010000000000001000080000500100000000000]  
[0000010000000000001000080000500100000000000]  
[0000010000000000001000080000500100000000000]
```

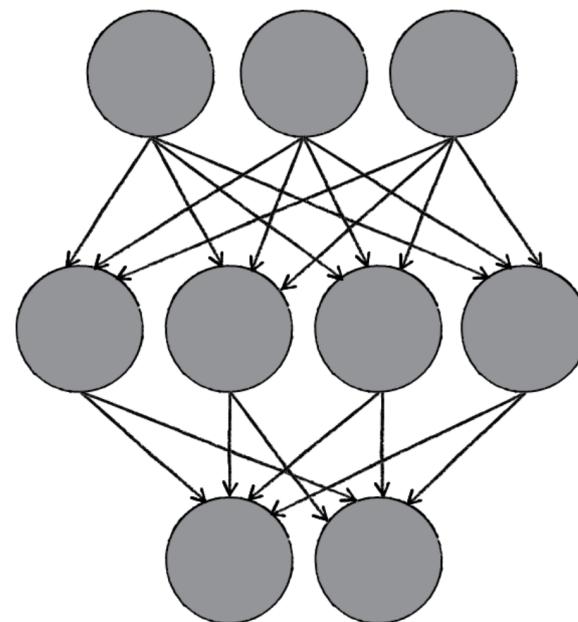


```
[00100010000000  
[00200030007002  
[3010000000900  
[005090010004000  
[60100001000008
```

```
[00000100000000000010000800005001000000000000]  
[00000100000000000010000800005001000000000000]  
[00000100000000000010000800005001000000000000]  
[00000100000000000010000800005001000000000000]  
[00000100000000000010000800005001000000000000]
```

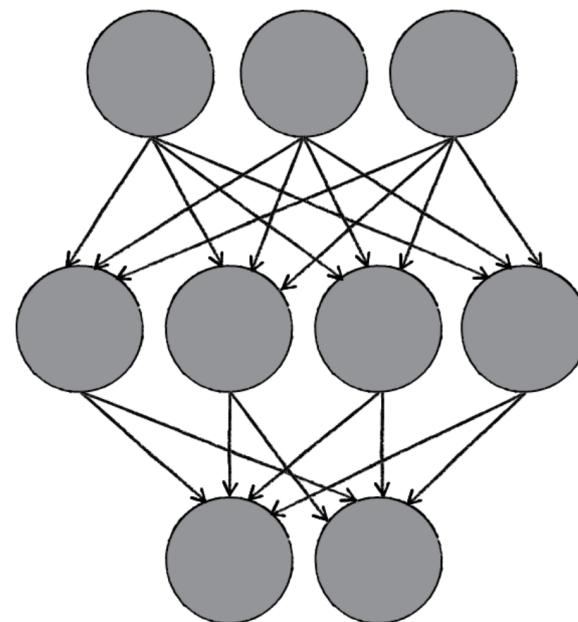
```
[ 001000010000000 ]  
[ 002000300070002 ]  
[ 301000000000900 ]  
[ 005090010004000 ]  
[ 601000010000008 ]
```

```
[00000100000000000010000800005001000000000000]  
[00000100000000000010000800005001000000000000]  
[00000100000000000010000800005001000000000000]  
[00000100000000000010000800005001000000000000]  
[00000100000000000010000800005001000000000000]
```



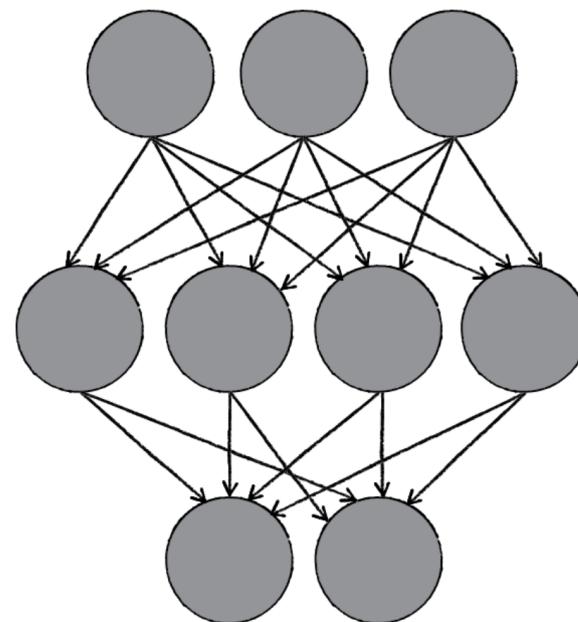
```
[001000010000000]  
[002000300070002]  
[301000000000900]  
[005090010004000]  
[601000010000008]
```

```
[00000100000000000010000800005001000000000000]  
[00000100000000000010000800005001000000000000]  
[00000100000000000010000800005001000000000000]  
[00000100000000000010000800005001000000000000]  
[00000100000000000010000800005001000000000000]
```



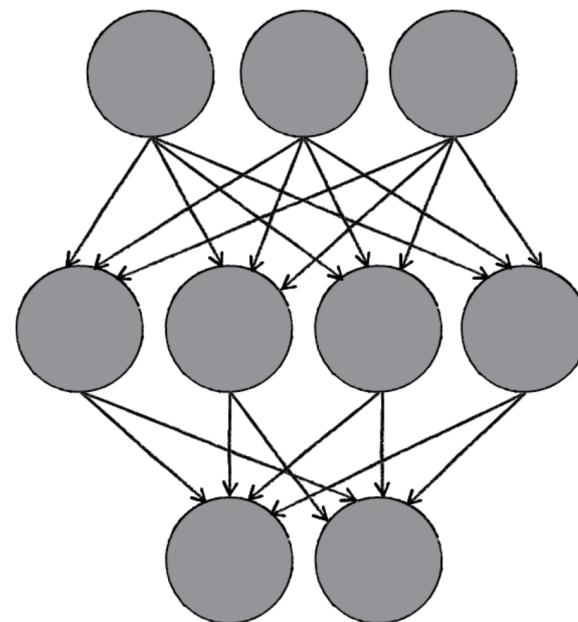
```
[ 001000010000000 ]  
[ 002000300070002 ]  
[ 301000000000900 ]  
[ 005090010004000 ]  
[ 601000010000008 ]
```

```
[00000100000000000010000800005001000000000000]  
[00000100000000000010000800005001000000000000]  
[00000100000000000010000800005001000000000000]  
[00000100000000000010000800005001000000000000]  
[00000100000000000010000800005001000000000000]
```



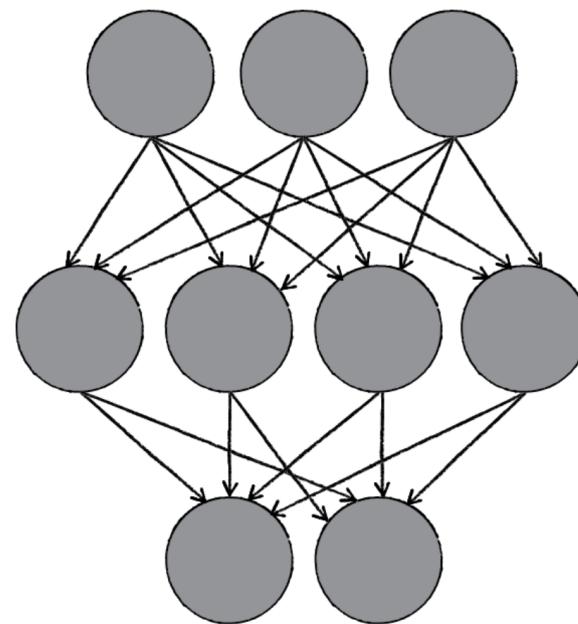
```
[001000010000000]  
[002000300070002]  
[301000000000900]  
[005090010004000]  
[601000010000008]
```

```
[00000100000000000010000800005001000000000000]  
[00000100000000000010000800005001000000000000]  
[00000100000000000010000800005001000000000000]  
[00000100000000000010000800005001000000000000]  
[00000100000000000010000800005001000000000000]
```



```
[001000010000000]  
[002000300070002]  
[301000000000900]  
[005090010004000]  
[601000010000008]
```

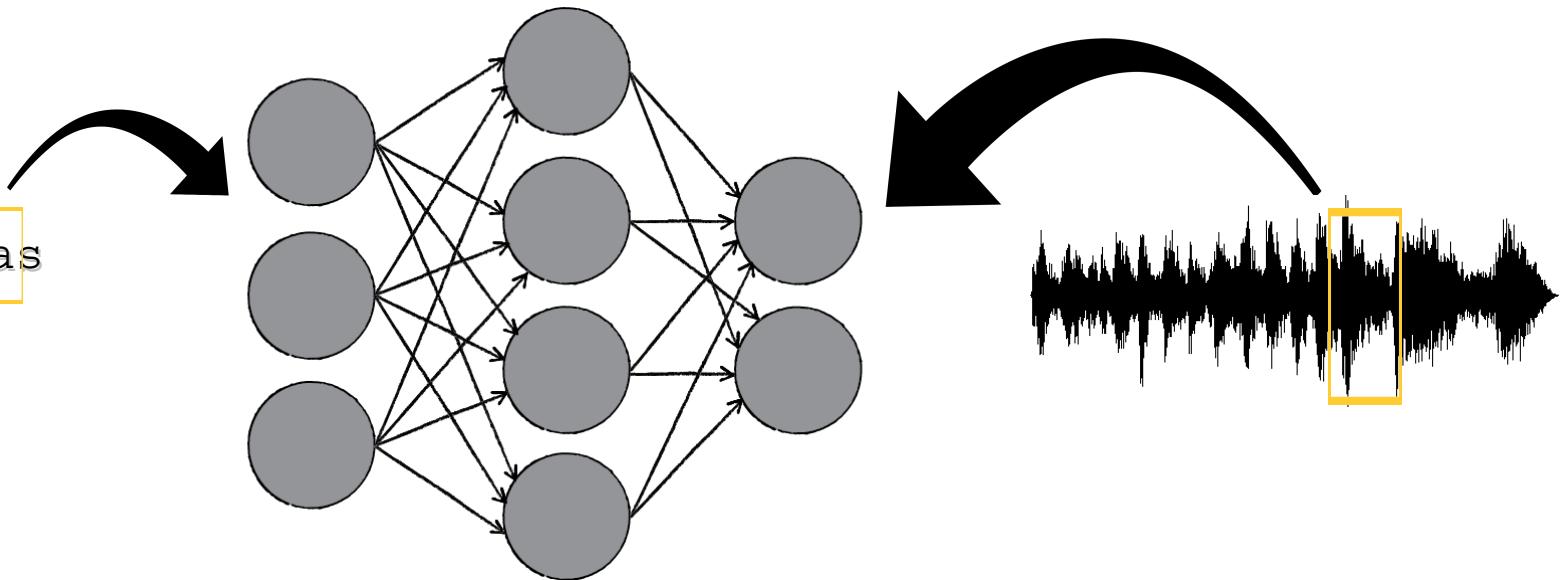
```
[00000100000000000010000800005001000000000000]  
[00000100000000000010000800005001000000000000]  
[00000100000000000010000800005001000000000000]  
[00000100000000000010000800005001000000000000]  
[00000100000000000010000800005001000000000000]
```



```
[001000010000000]  
[002000300070002]  
[301000000000900]  
[005090010004000]  
[601000010000008]
```



Moscow always has
traffic jams.



References / Resources

- CSTR slides: speech.zone
- Ossian: github.com/CSTR-Edinburgh/Ossian
- Merlin: [github.com/CSTR-Edinburgh](https://github.com/CSTR-Edinburgh/Merlin)