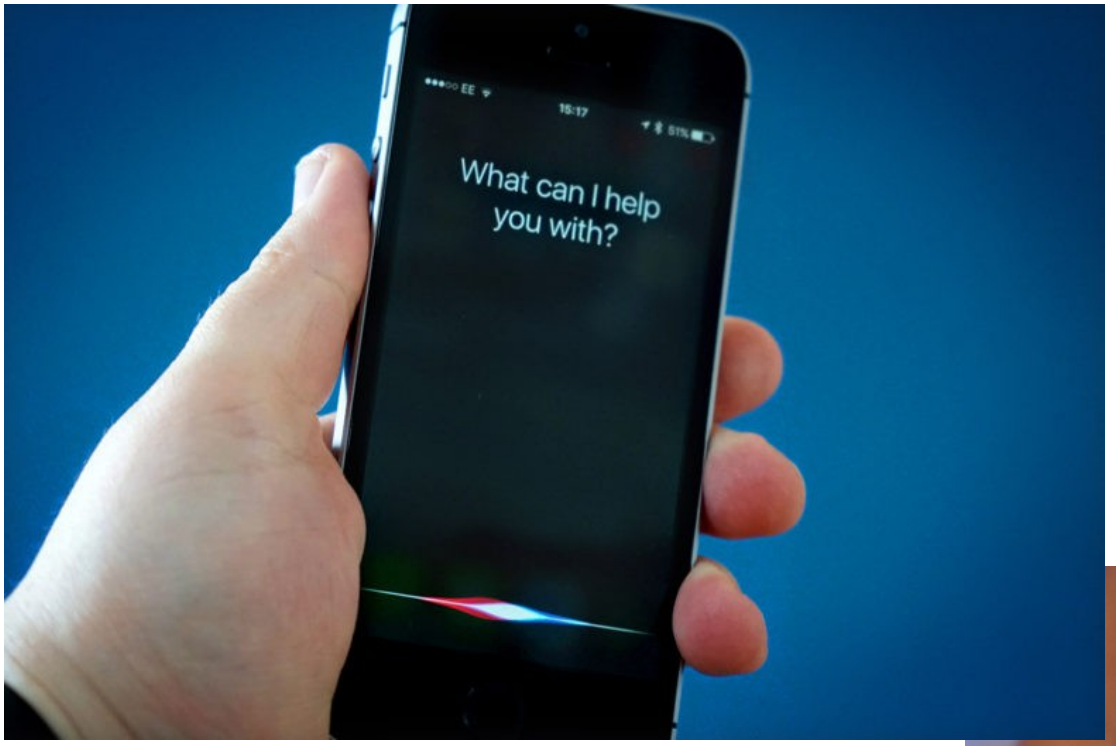
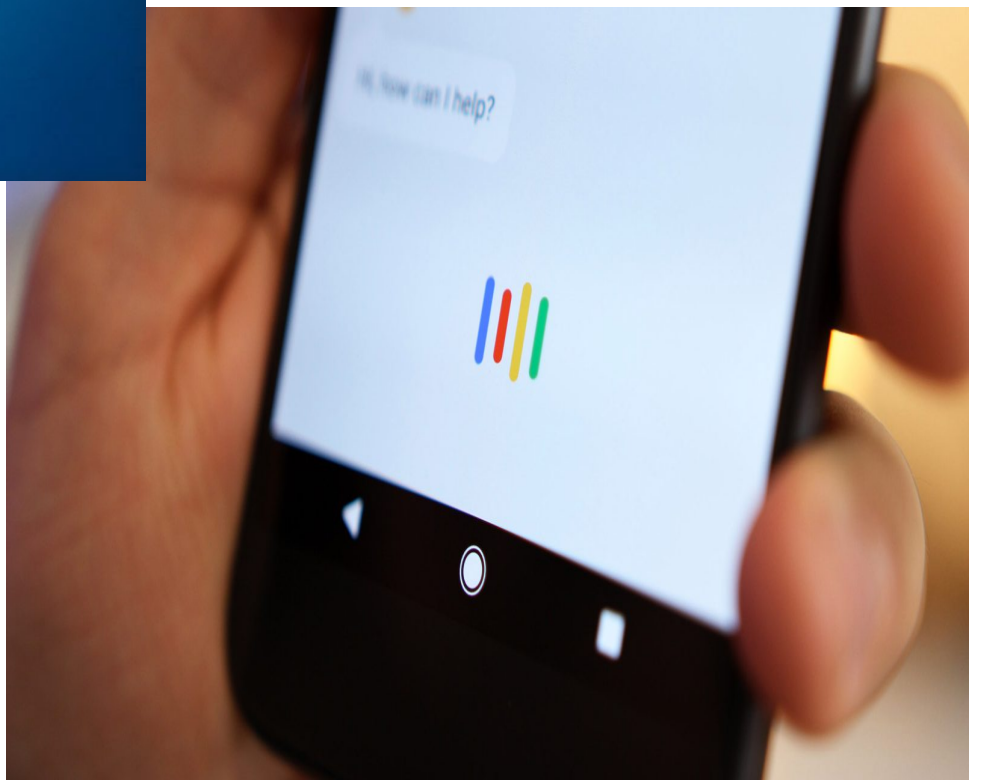
Speech Synthesis Workshop
@ Higher School of Economics
Moscow

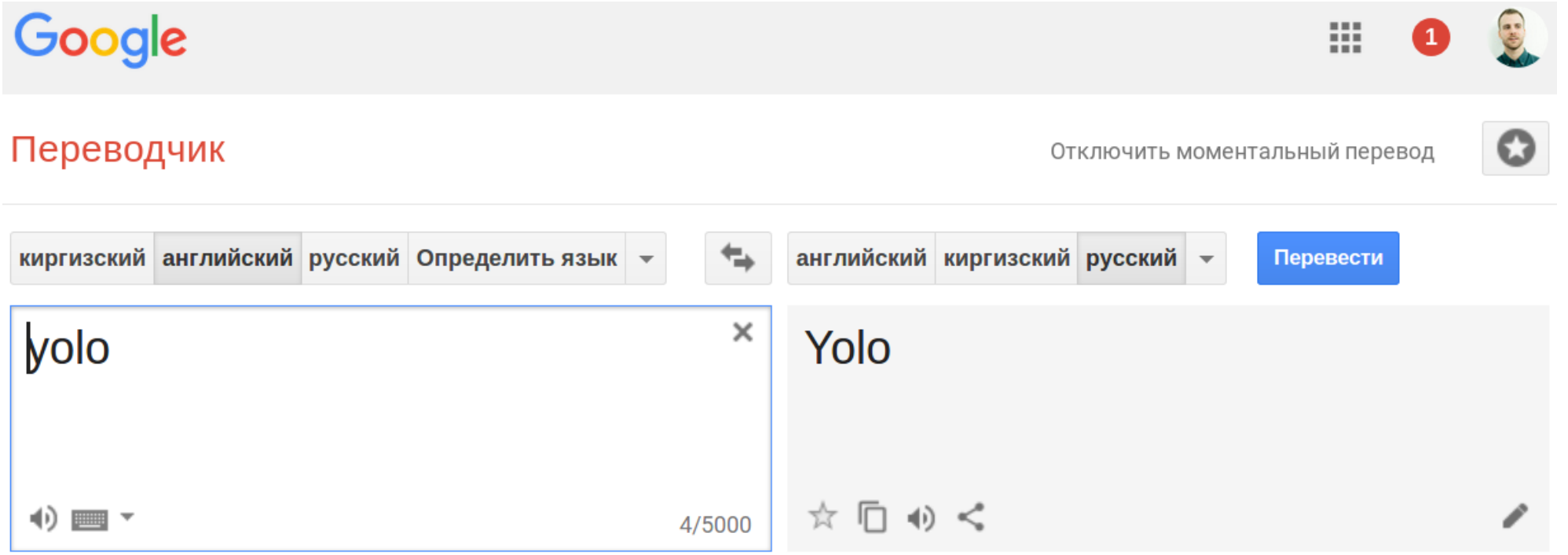
Josh Meyer

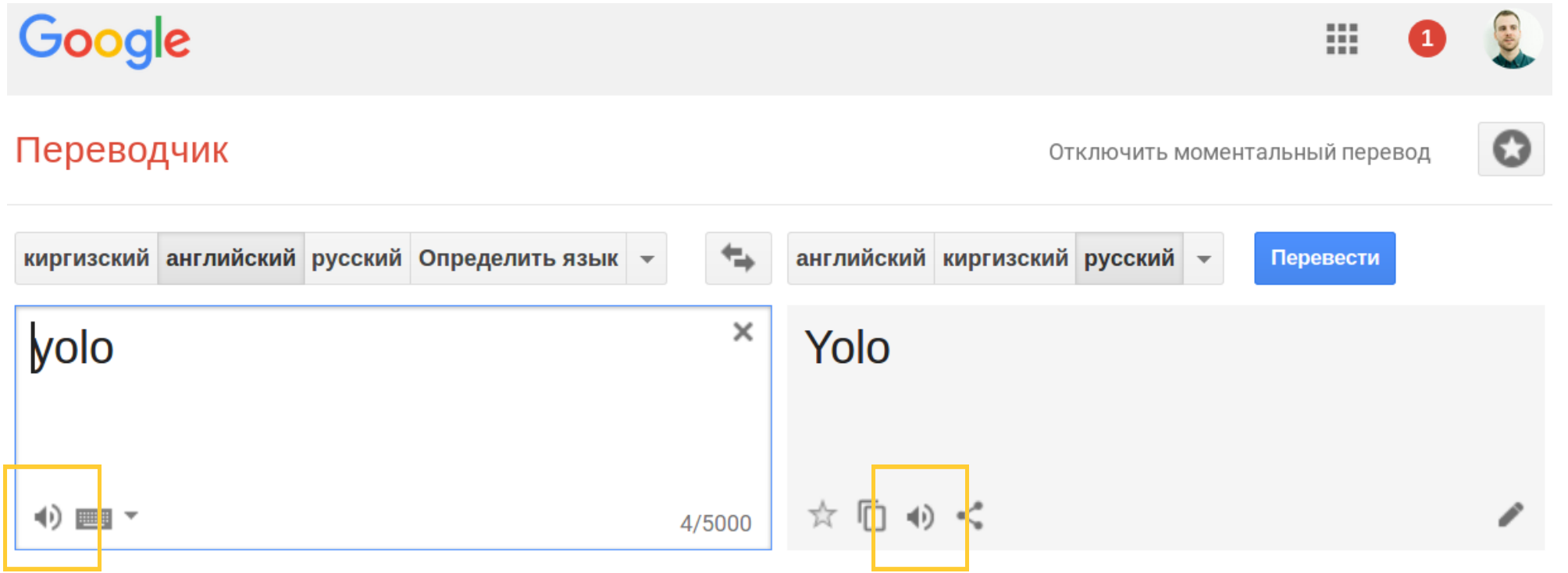
@joshmeyerphd

jrmeyer.github.io

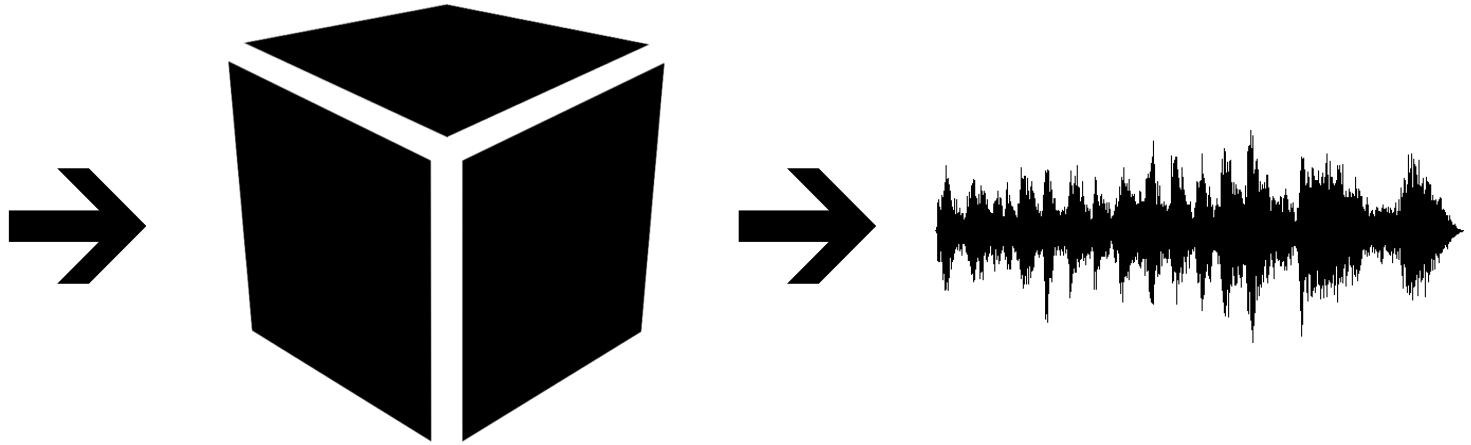

What can I help
you with?




Google
1
Переводчик
Отключить моментальный перевод
киргизский
английский
русский
Определить язык
английский
киргизский
русский
Перевести
yolo
4/5000
Yolo


Google
1
Переводчик
Отключить моментальный перевод
киргизский
английский
русский
Определить язык
английский
киргизский
русский
Перевести
yolo
4/5000
Yolo

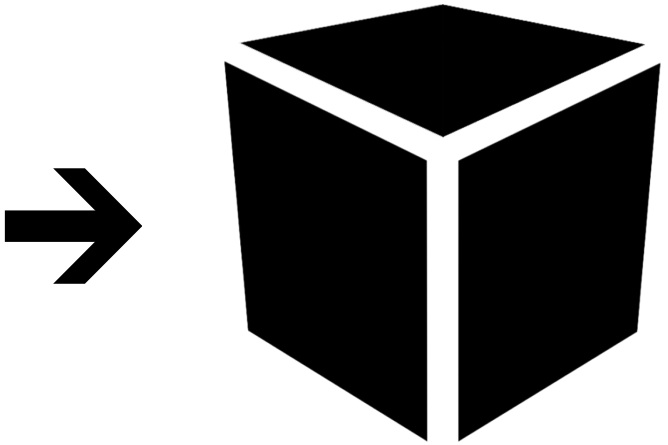
Moscow always has traffic jams.



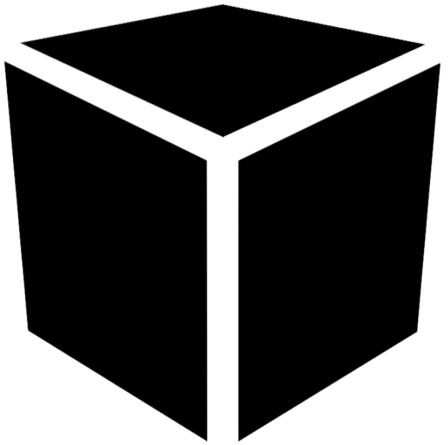
RAW TEXT

RAW AUDIO

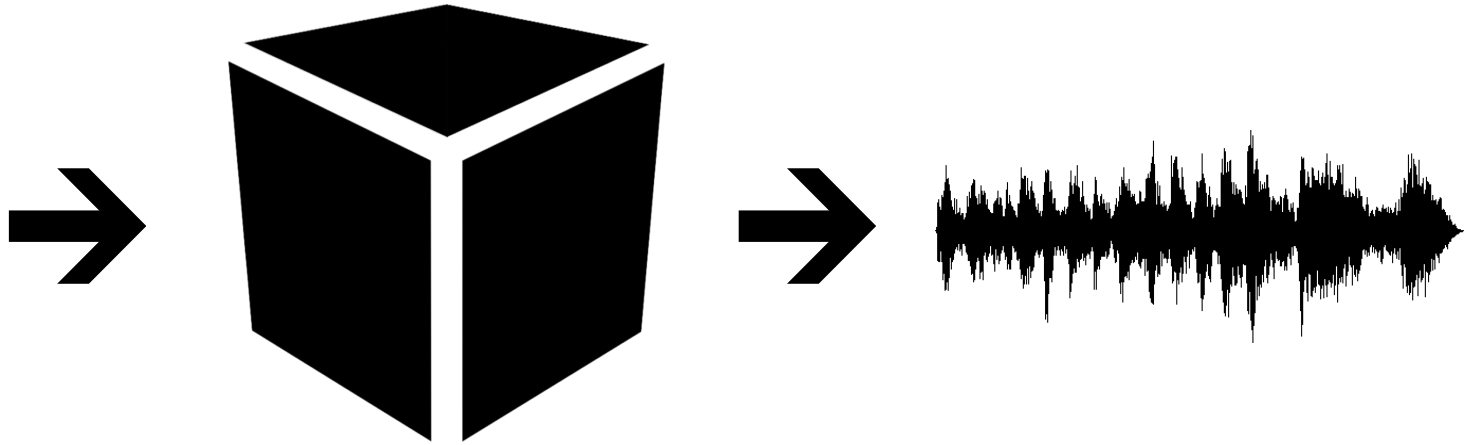
Moscow always has traffic jams.



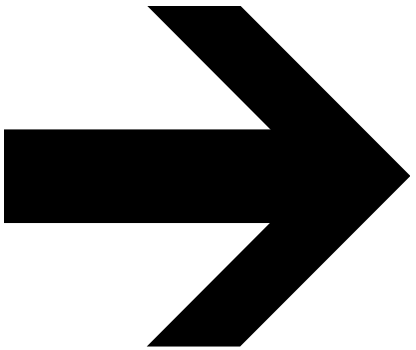
RAW TEXT → **SPEECH SYNTHESIZER** → **RAW AUDIO**

Moscow always has traffic jams. →  → 

Moscow always has traffic jams.

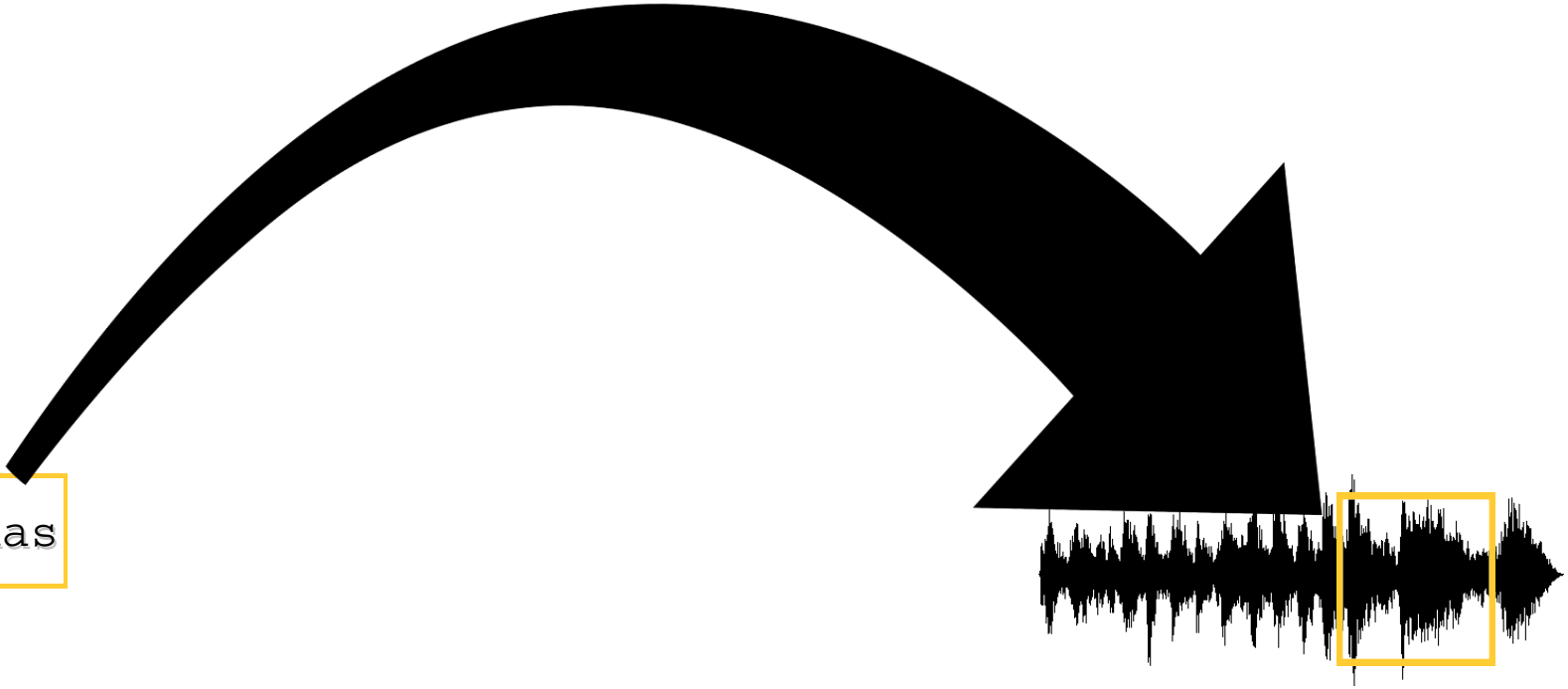


Moscow always has traffic jams.



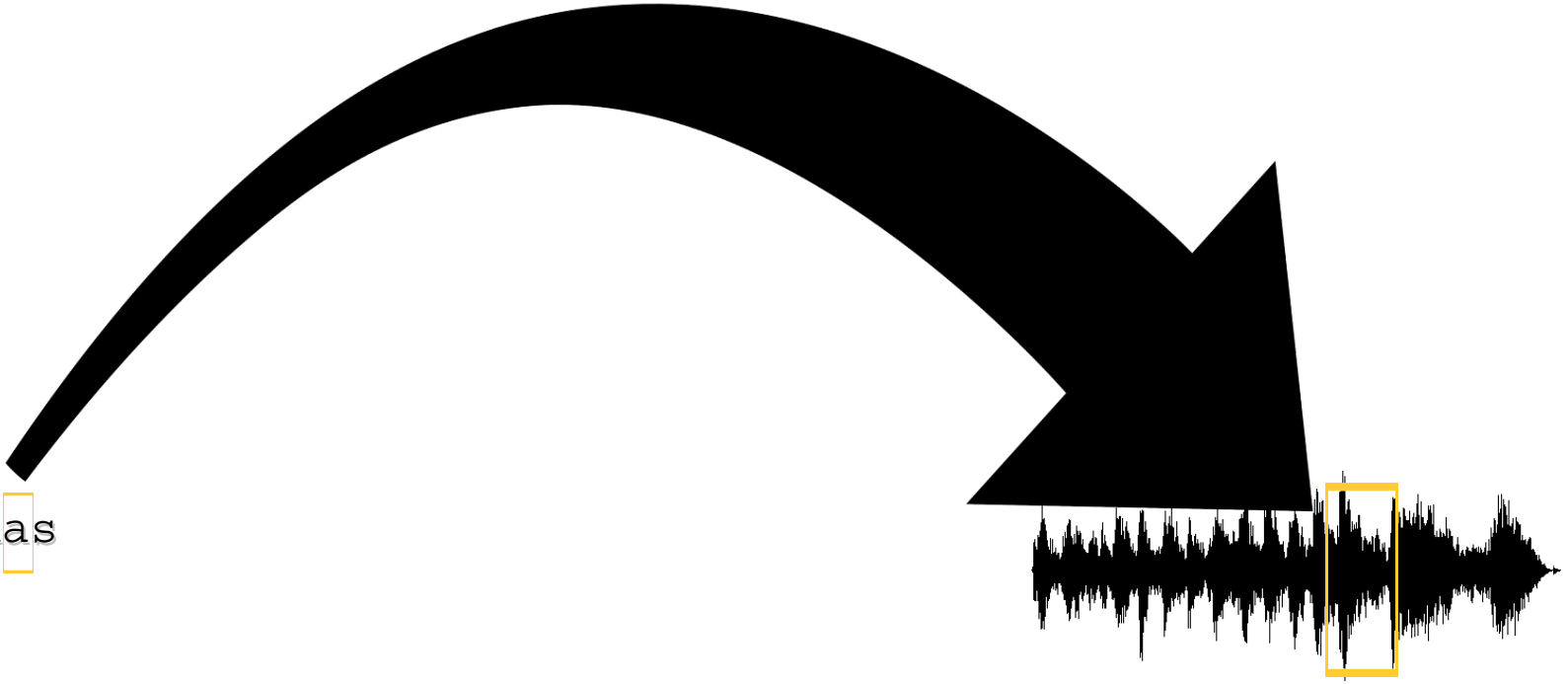
HARD

Moscow always has
traffic jams.



HARD

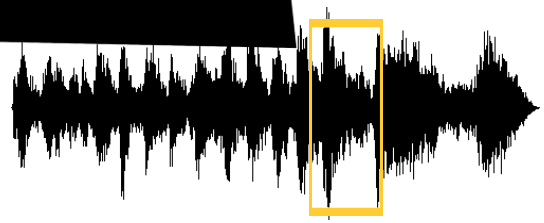
Moscow always has
traffic jams.



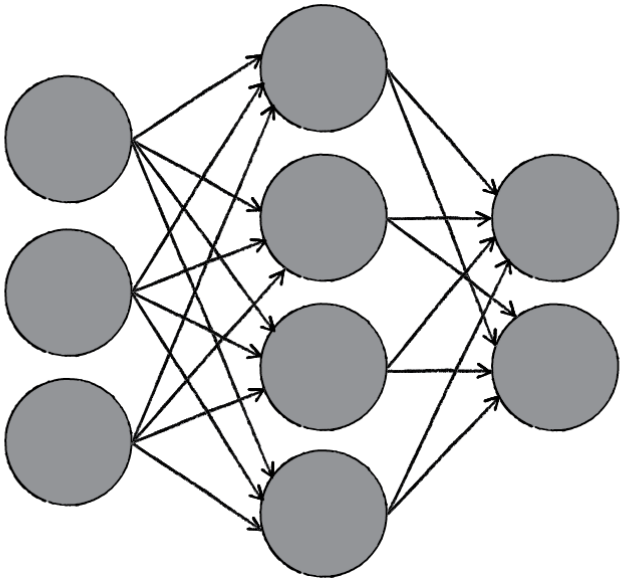
EASIER

REGRESSION

Moscow always has
traffic jams.



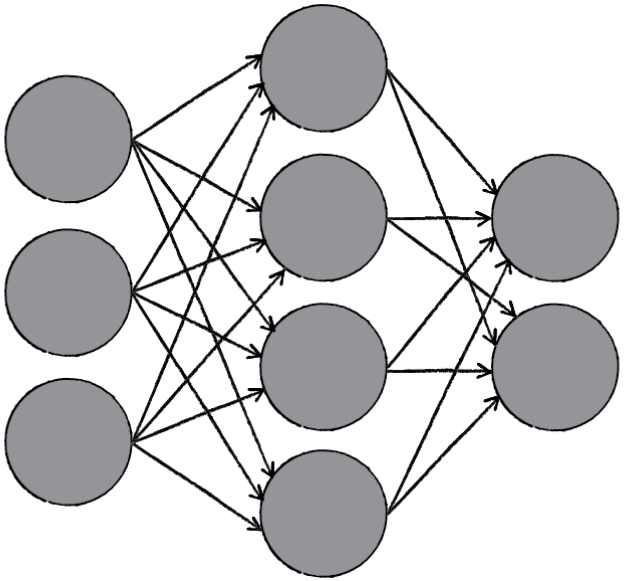
Moscow always has traffic jams.

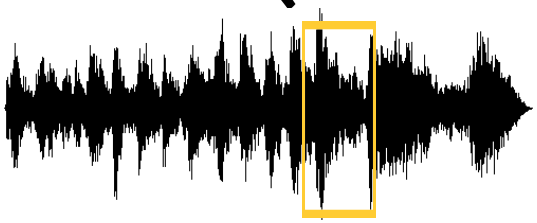


BECAUSE

DNNs

Moscow always has
traffic jams.





Moscow always has
traffic jams.

[0010000007000230100000000009000]

[6010000000000008]

FRONTEND

Moscow always has
traffic jams.

[0010000007000230100000000009000]

[6010000000000008]

BACKEND

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]

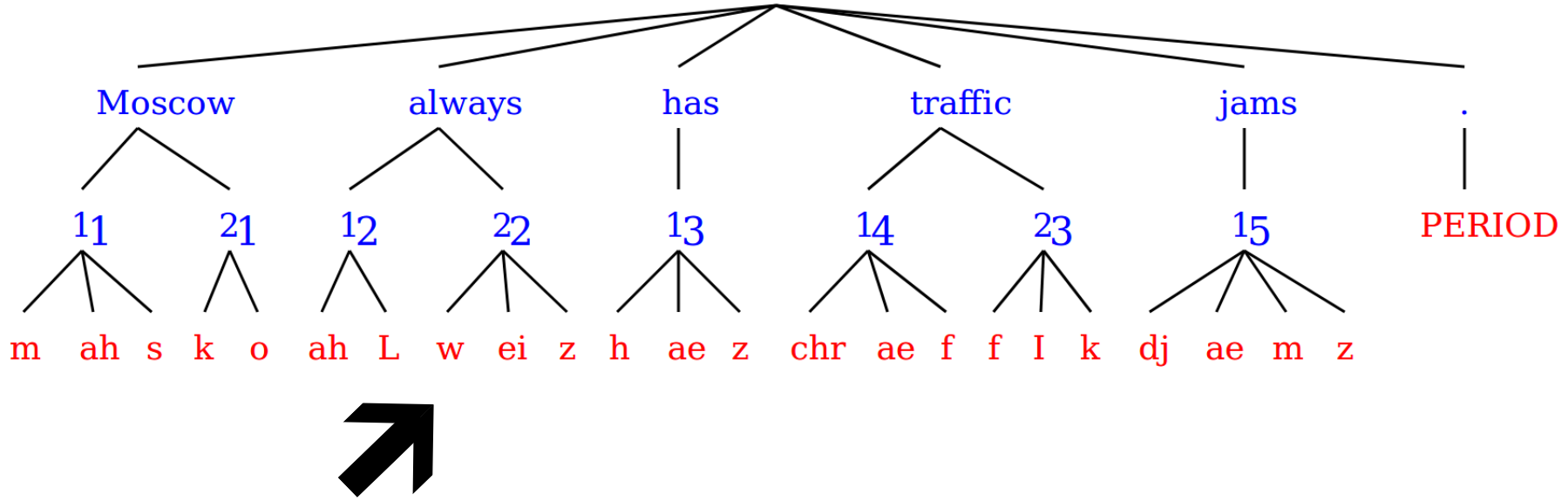
PHONEMES + WORDS + POS

Moscow always has
traffic jams.

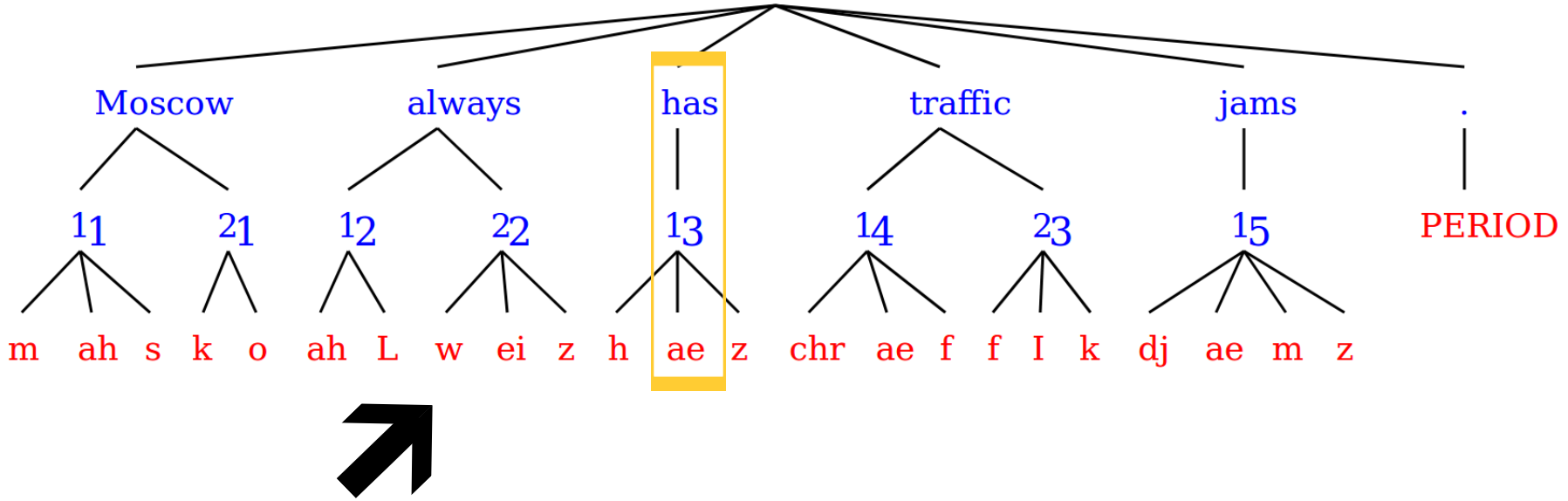
[NOUN [M ah][s k o]] [ADVERB [ah L][w ei z]]
[VERB [h ae z]] [NOUN [chr ae][f I k]]
[NOUN [dj ae m z]]

PHONEMES + WORDS + POS
+ SYLLABLES

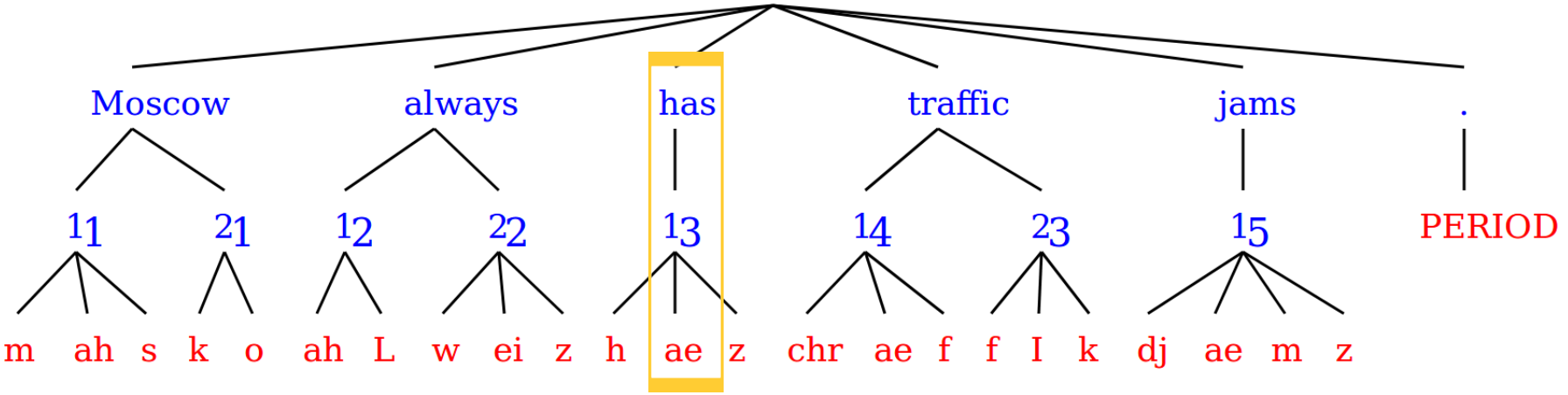
Moscow always has
traffic jams.



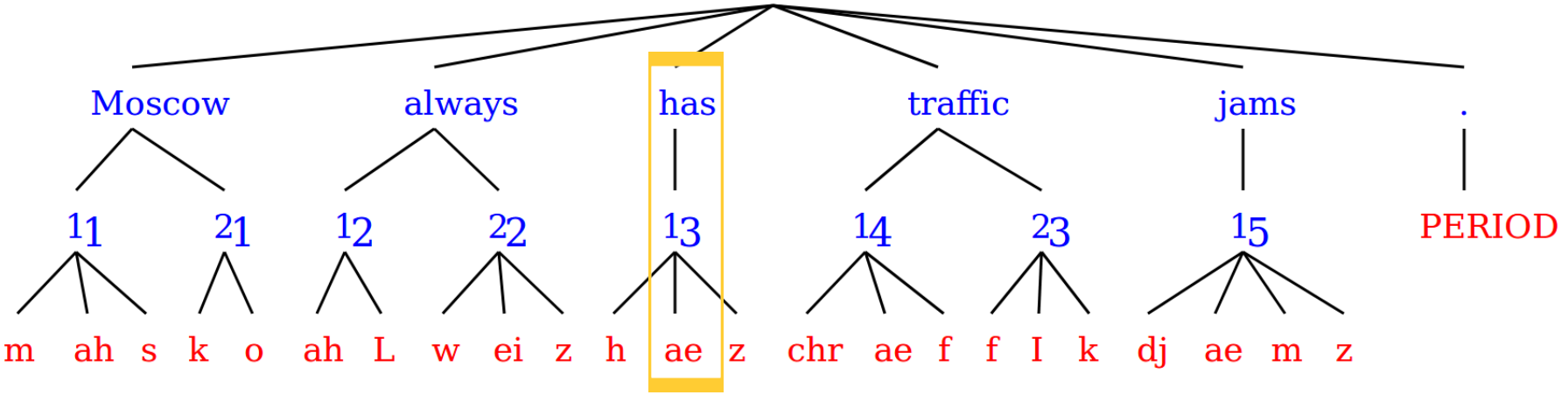
Moscow always has
traffic jams.



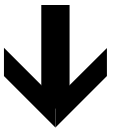
Moscow always has
traffic jams.



i+/0:_NA_/1:sil/2:ah/3:f/4:y/5:0/6:8/7:4/8:1/9:8/10:0/11:9/12:0/13:0/14:76/15:0/16:1/17:76/18:1/19:9/20:2/21:1/22:1/23:78/24:9/25:1/26:79/27:9/28:1/29:0.630101876438/30:0.685195227915/31:0.895005526298/32:0.016938420738 6/33:0.0696863228001/34:0.184964316317/35:0.053507651946 5/36:0.030208759648 4/37:-0.148247666767/38:0.011018346386 2/39:0.006837177869 86/40:0.103955351166/41:0.018273938283 6/[6]

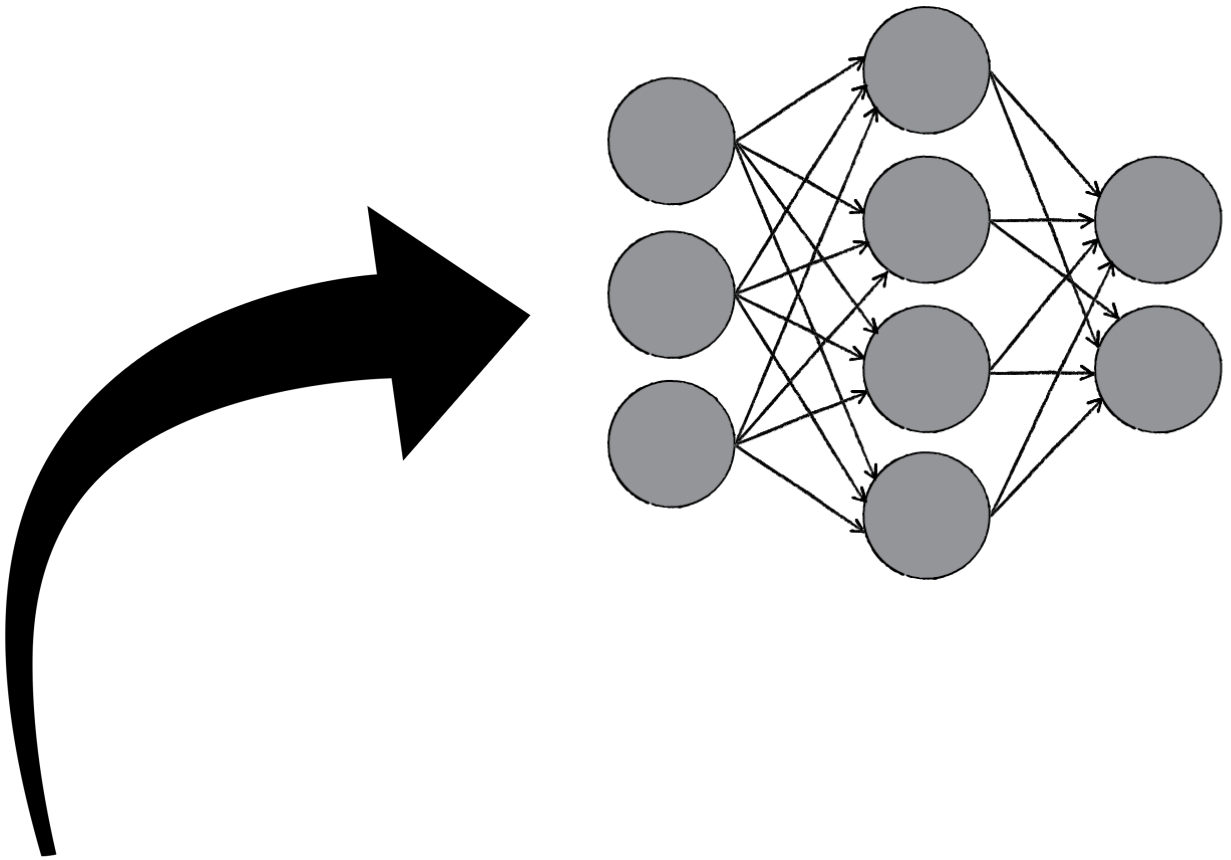


i+/0:_NA_/1:sil/2:ah/3:f/4:y/5:0/6:8/7:4/8:1/9:8/10:0/11:9/12:0/13:0/14:76/15:0/16:1/17:76/18:1/19:9/20:2/21:1/22:1/23:78/24:9/25:1/26:79/27:9/28:1/29:0.630101876438/30:0.685195227915/31:0.895005526298/32:0.016938420738 6/33:0.0696863228001/34:0.184964316317/35:0.053507651946 5/36:0.030208759648 4/37:-0.148247666767/38:0.011018346386 2/39:0.006837177869 86/40:0.103955351166/41:0.018273938283 6/[6]

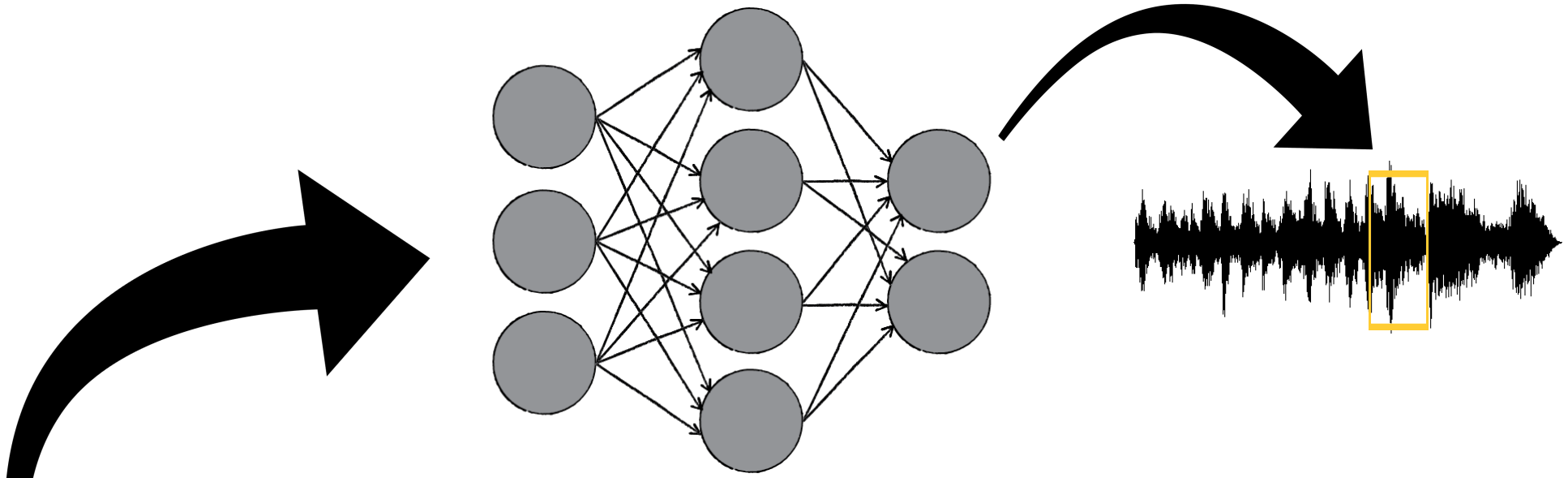


[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0]

[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0]



[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 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 0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0]

THE FRONTEND

1) MOTIVATION

2) TEXT → XML

Фильмсем 13 е 13:30 сехетре пуçланĕç.

Фильмсем 13 е 13:30 сехетре пуçланӗç.

<utt text="Фильмсем 13 е 13:30 сехетре пуçланӗç."
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="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="пуçланӗç" token_class="word"> </token>
<token text="." token_class="punctuation"> </token>
<token text="_END_" token_class="_END_"> </token>

</utt>

Фильмсем 13 е 13:30 сехетре пуçланӗç.

UTTERANCE → <utt text="Фильмсем 13 е 13:30 сехетре пуçланӗç."
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="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="пуçланӗç" token_class="word"> </token>
<token text="." token_class="punctuation"> </token>
<token text="_END_" token_class="_END_"> </token>

</utt>

Фильмсем 13 е 13:30 сехетре пуçланӗç.

UTTERANCE → <utt text="Фильмсем 13 е 13:30 сехетре пуçланӗç."
waveform="/home/ubuntu/Ossian/corpus/chv/speakers/news/wa
v/17448-0006.wav" utterance_name="17448-0006"
PROCESS → 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="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="пуçланӗç" token_class="word"> </token>
<token text="." token_class="punctuation"> </token>
<token text="_END_" token_class="_END_"> </token>

</utt>

Фильмсем 13 е 13:30 сехетре пуçланӗç.

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

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="пуçланӗç" token_class="word"> </token>
<token text="." token_class="punctuation"> </token>
<token text="_END_" token_class="_END_"> </token>

</utt>

Фильмсем 13 е 13:30 сехетре пуçланӗç.

<utt text="Фильмсем 13 е 13:30 сехетре пуçланӗç."
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="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="пуçланӗç" token_class="word"> </token>
<token text="." token_class="punctuation"> </token>
<token text="_END_" token_class="_END_"> </token>

</utt>

Фильмсем 13 е 13:30 сехетре пуçланĕç.

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

safetext="_CYRILLICCAPITALLETTEREF__CYRILLICSMALLLETTERI__
CYRILLICSMALLLETTEREL__CYRILLICSMALLLETTERSOFTSIGN__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="0.00063074819014"
vsm_d9="0.0073482650453 2" vsm_d10="0.0093897125 4099"
>

</token>

Фильмсем 13 е 13:30 сехетре пуçланĕç.

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

SAFETXT → safetext="_CYRILLICCAPITALLETTEREF__CYRILLICSMALLLETTERI__CYRILLICSMALLLETTEREL__CYRILLICSMALLLETTERSOFTSIGN__CYRILLICSMALLLETTEREM__CYRILLICSMALLLETTERES__CYRILLICSMALLLETTERIE__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="0.00063074819014"
vsm_d9="0.0073482650453 2" vsm_d10="0.0093897125409 9"
>

</token>

Фильмсем 13 е 13:30 сехетре пуçланĕç.

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

SAFETXT → safetext="_CYRILLICCAPITALLETTEREF__CYRILLICSMALLLETTERI__CYRILLICSMALLLETTEREL__CYRILLICSMALLLETTERSOFTSIGN__CYRILLICSMALLLETTEREM__CYRILLICSMALLLETTERES__CYRILLICSMALLLETTERIE__CYRILLICSMALLLETTEREM_"

VECTORS → 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="0.00063074819014" vsm_d9="0.0073482650453 2" vsm_d10="0.0093897125409 9"
>

</token>

Фильмсем 13 е 13:30 сехетре пуçланĕç.

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

SAFETXT → safetext="_CYRILLICCAPITALLETTEREF__CYRILLICSMALLLETTERI__CYRILLICSMALLLETTEREL__CYRILLICSMALLLETTERSOFTSIGN__CYRILLICSMALLLETTEREM__CYRILLICSMALLLETTERES__CYRILLICSMALLLETTERIE__CYRILLICSMALLLETTEREM_"

VECTORS → 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="0.00063074819014" vsm_d9="0.0073482650453 2" vsm_d10="0.0093897125409 9"
>

PHONES →

</token>

Фильмсем 13 е 13:30 сехетре пуçланĕç.

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

SAFETXT → safetext="_DIGITONE__DIGITTHREE_"

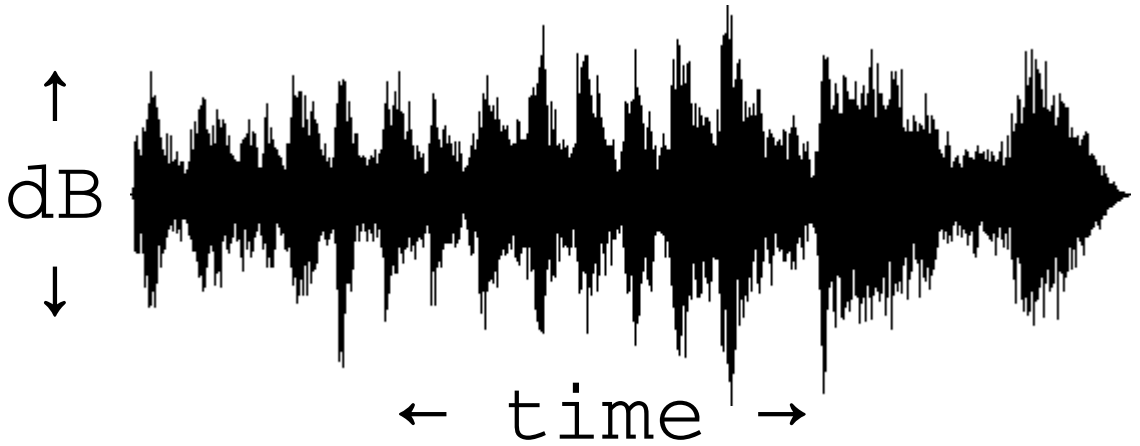
VECTORS → vsm_d1="0.175666404262" vsm_d2="0.152223466799"
vsm_d3="-0.0307015100812" vsm_d4="0.099961494321 1"
vsm_d5="0.0698700498727" vsm_d6="0.25982222856"
vsm_d7="0.0213851732067" vsm_d8="-0.00852679385956"
vsm_d9="-0.0024125017034" vsm_d10="0.0225874466051"
>

PHONES →

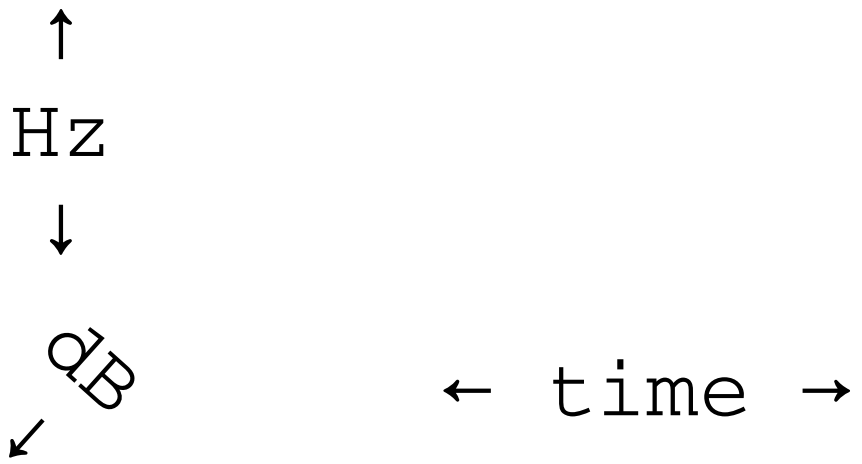
</token>

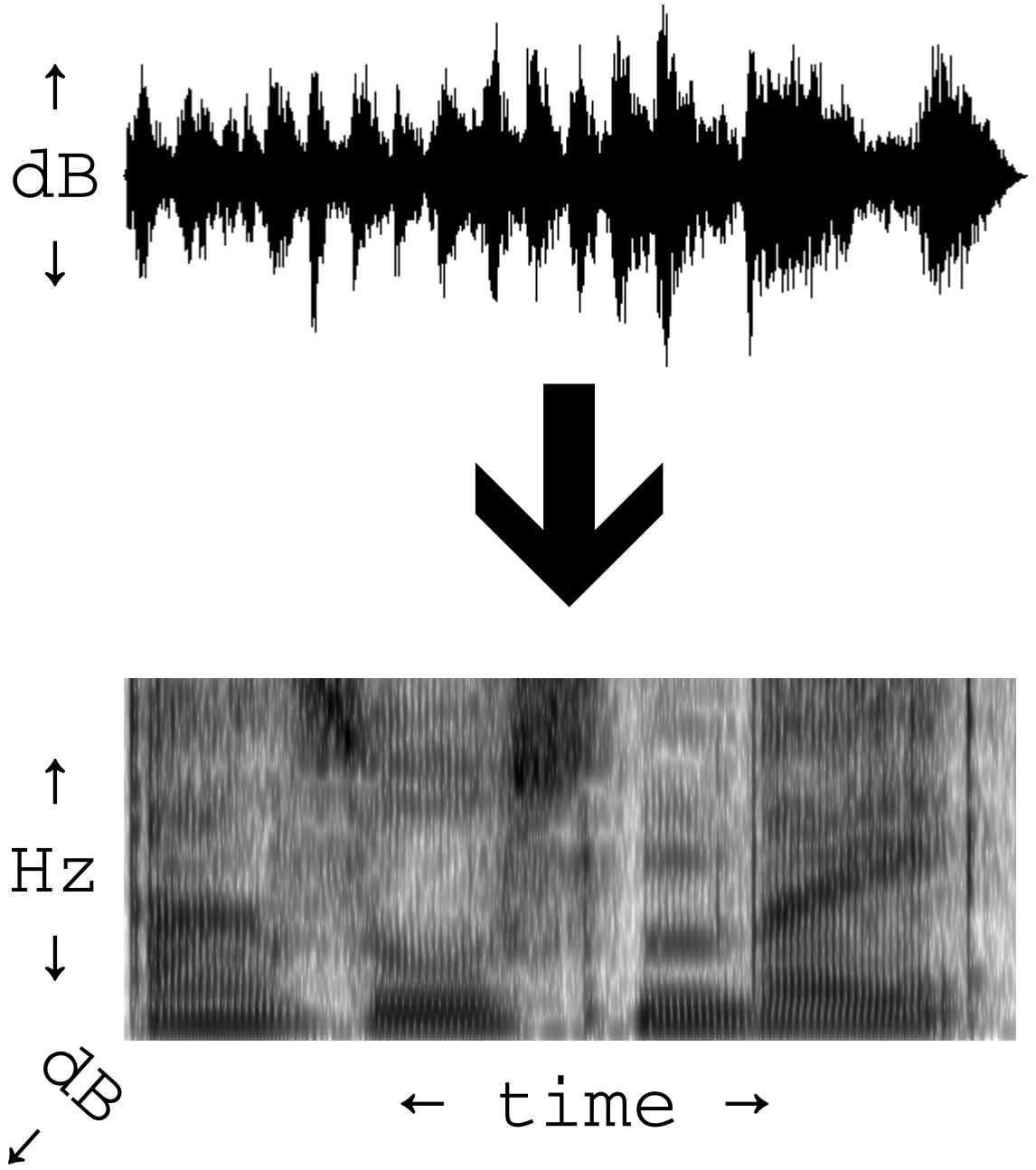
THE BACKEND

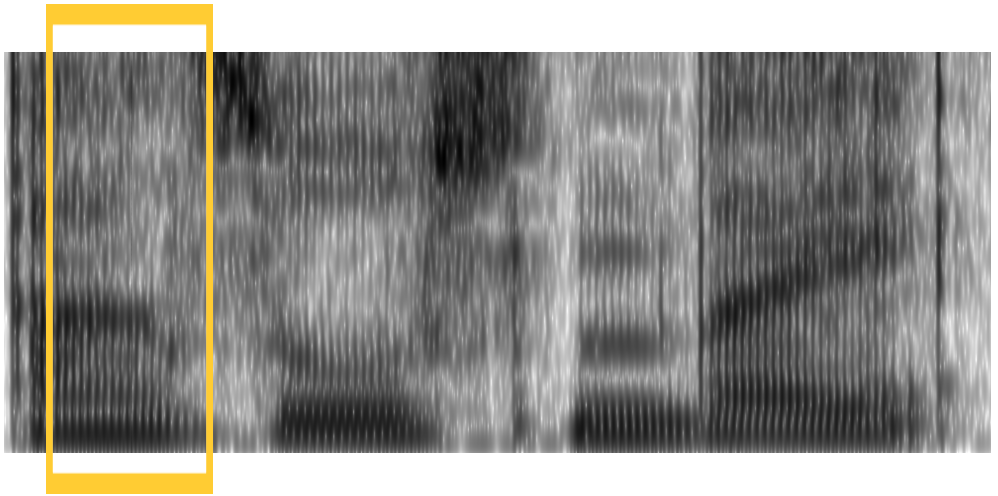
(audio feature extraction)

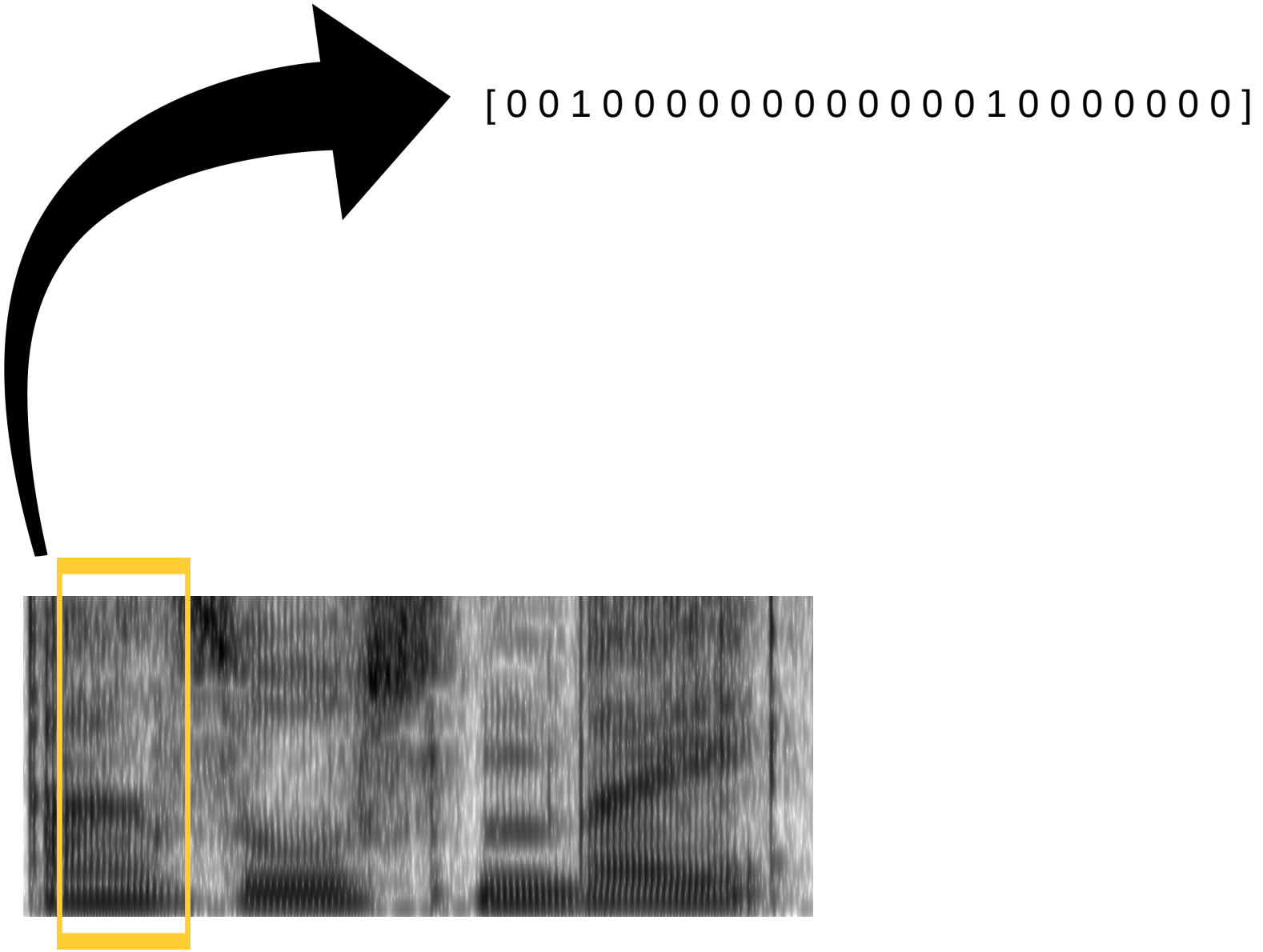

↑
dB
↓
← time →


↑
dB
↓


↑
Hz
↓
dB
↙
← time →


↑
dB
↓
↑
Hz
↓
dB
← time →




[0 0 1 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0]

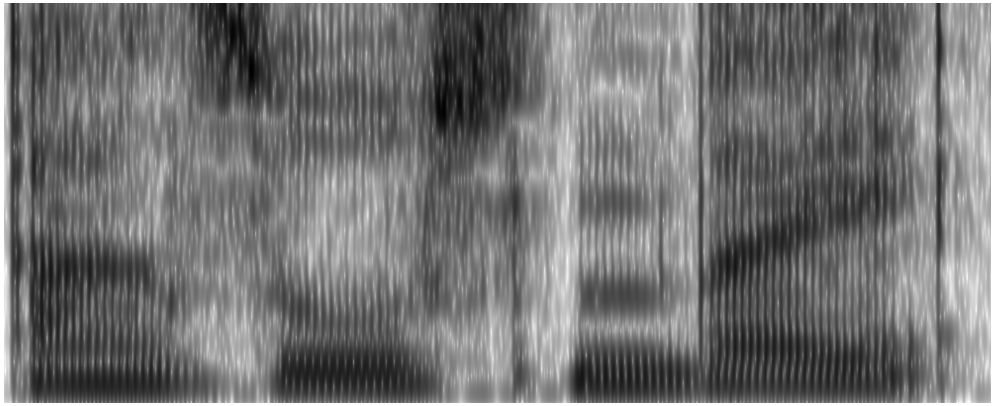
REGRESSION

REGRESSION

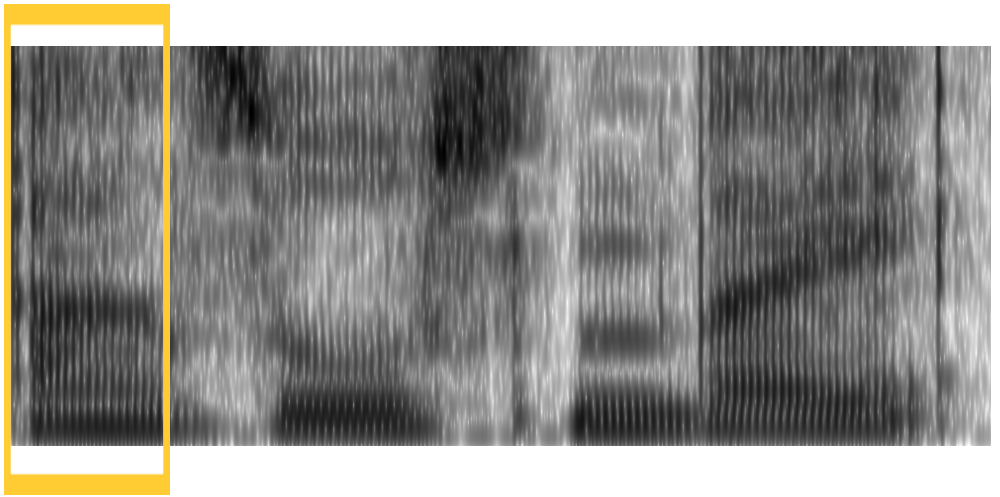
1) LABELED DATA

Moscow always has traffic jams.

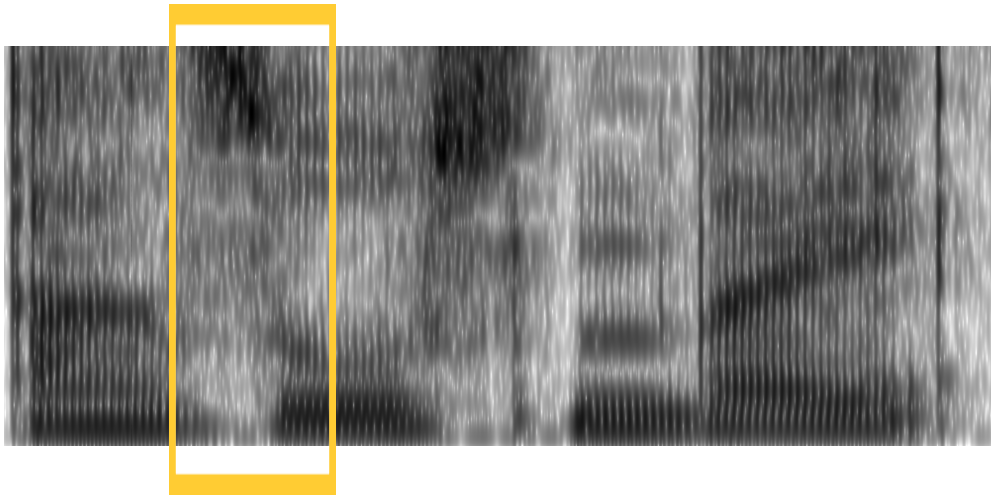
Moscow always has traffic jams.



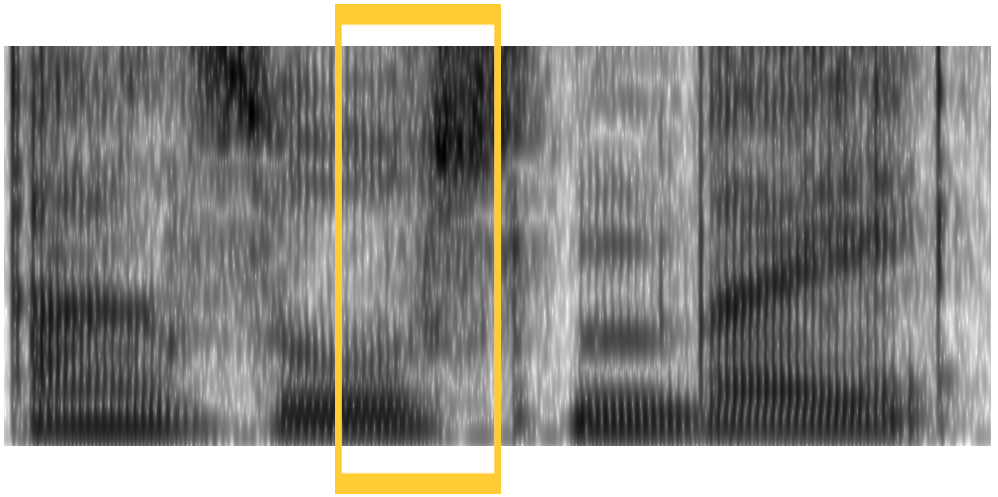
Moscow always has traffic jams.



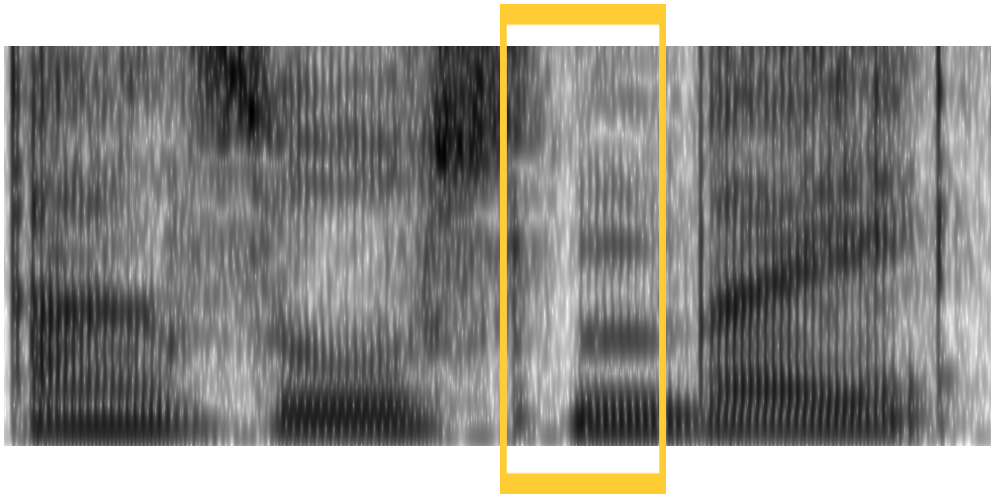
Moscow always has traffic jams.



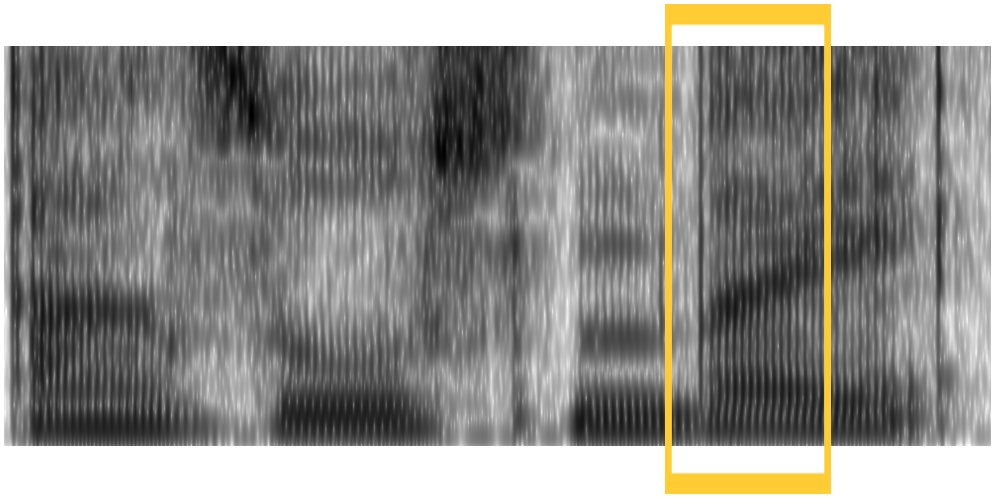
Moscow always has traffic jams.



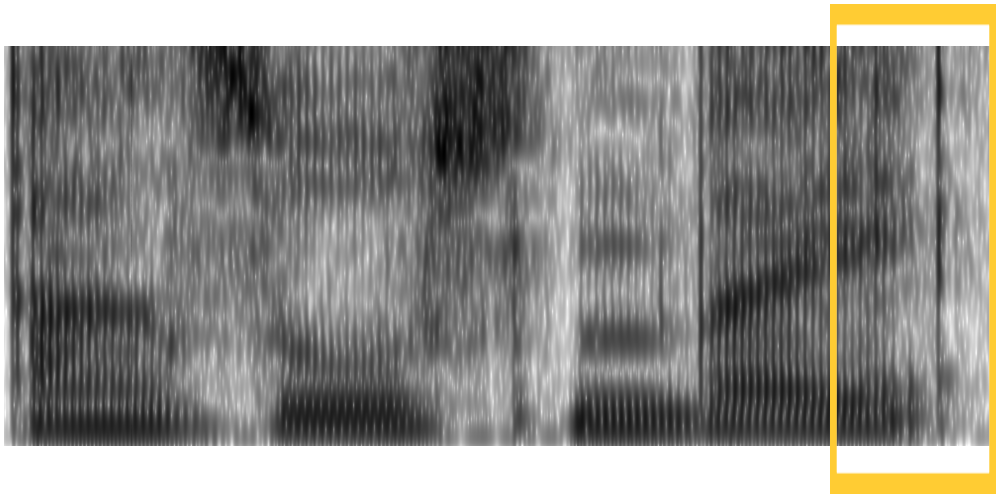
Moscow always has traffic jams.



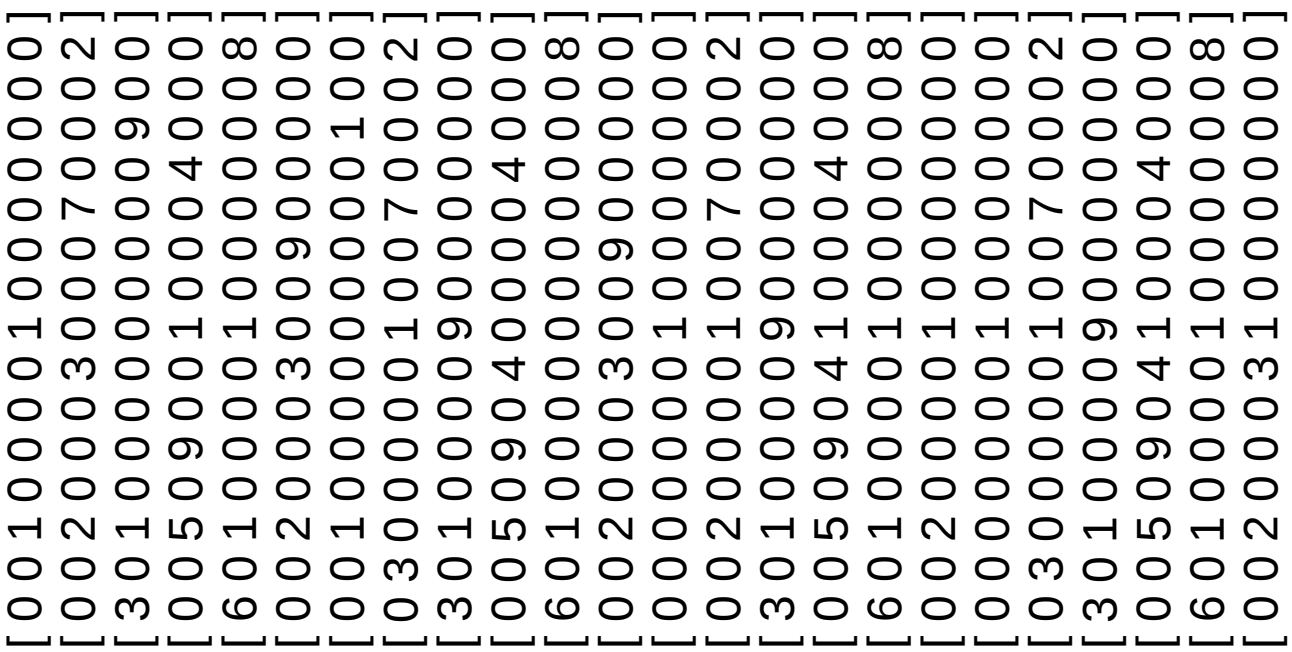
Moscow always has traffic jams.

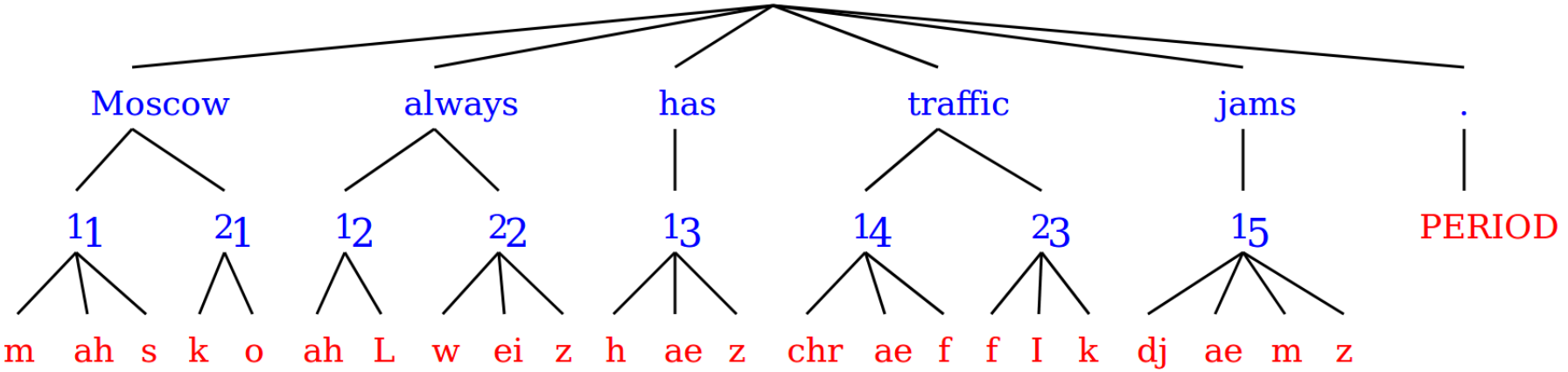


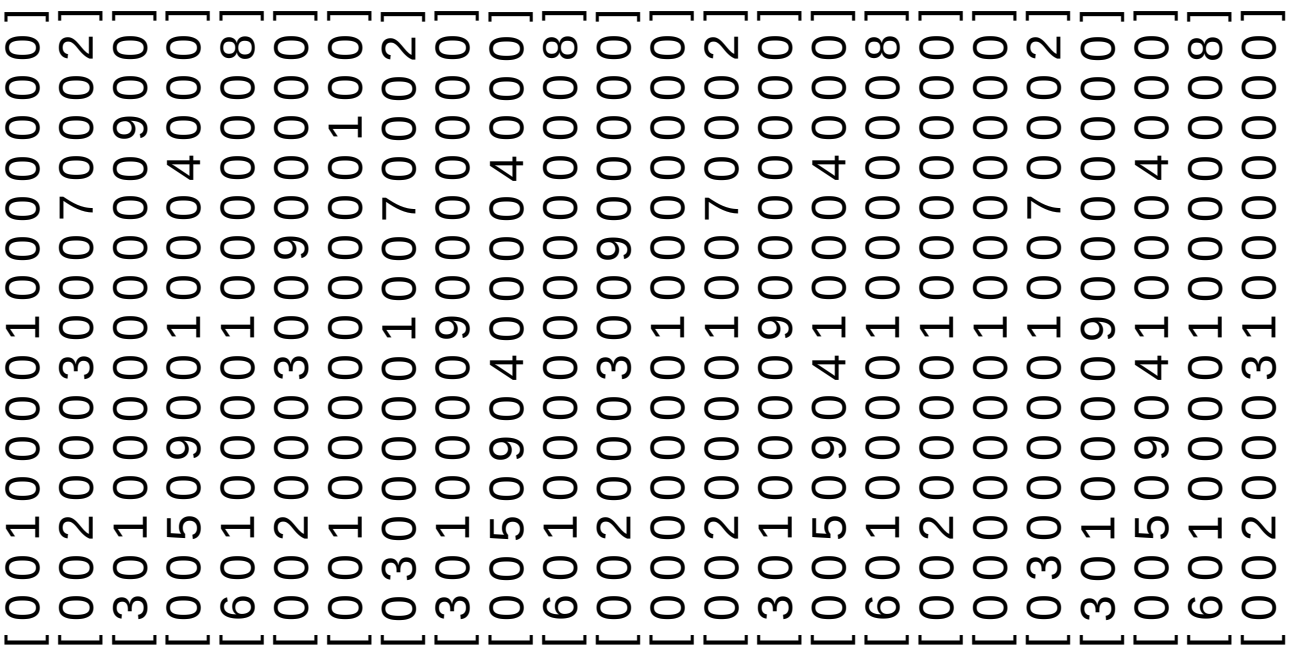
Moscow always has traffic jams.



Moscow always has traffic jams.




Moscow
always
has
traffic
jams
.
11
21
12
22
13
14
23
15
PERIOD
m ah s k o ah L w ei z h ae z chr ae f f I k dj ae m z



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/
18:1/19:9/20:2/21:1/22:1/23:78/24:9/25:1/26:79/27:9/28:1/29:0.630101876438/30:0.6851
95227915/31:0.895005526298/32:0.0169384207386/33:0.0696863228001/34:0.18496

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/
18:1/19:9/20:2/21:1/22:1/23:78/24:9/25:1/26:79/27:9/28:1/29:0.630101876438/30:0.6851
95227915/31:0.895005526298/32:0.0169384207386/33:0.0696863228001/34:0.18496

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/
18:1/19:9/20:2/21:1/22:1/23:78/24:9/25:1/26:79/27:9/28:1/29:0.630101876438/30:0.6851
95227915/31:0.895005526298/32:0.0169384207386/33:0.0696863228001/34:0.18496

.

.

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

[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0 0]

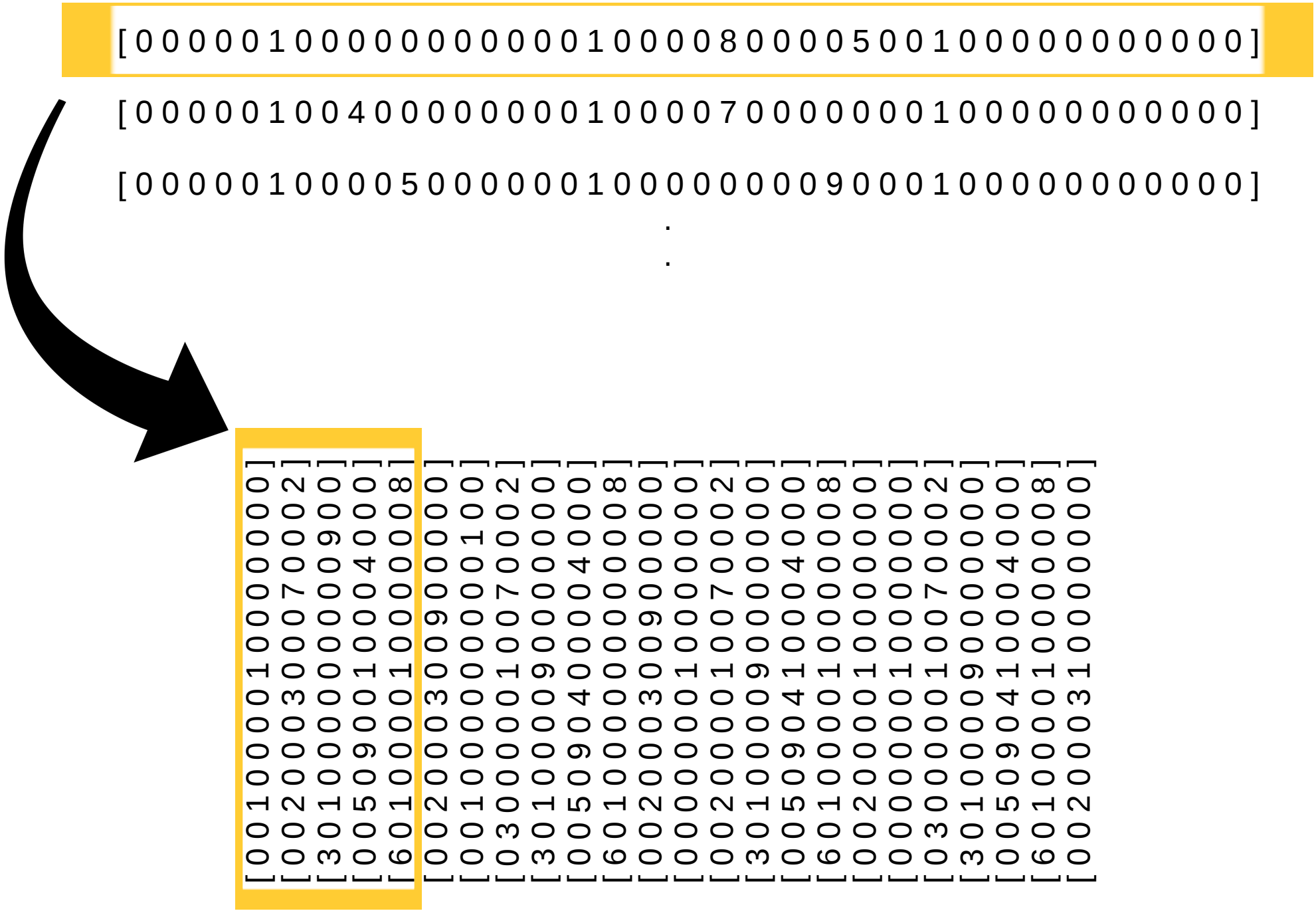
[0 0 0 0 0 1 0 0 4 0 0 0 0 0 0 0 0 1 0 0 0 0 7 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0]

[0 0 0 0 0 1 0 0 0 0 5 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 9 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0]

.

.

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


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

[0000010000000000010000800005001000000000]

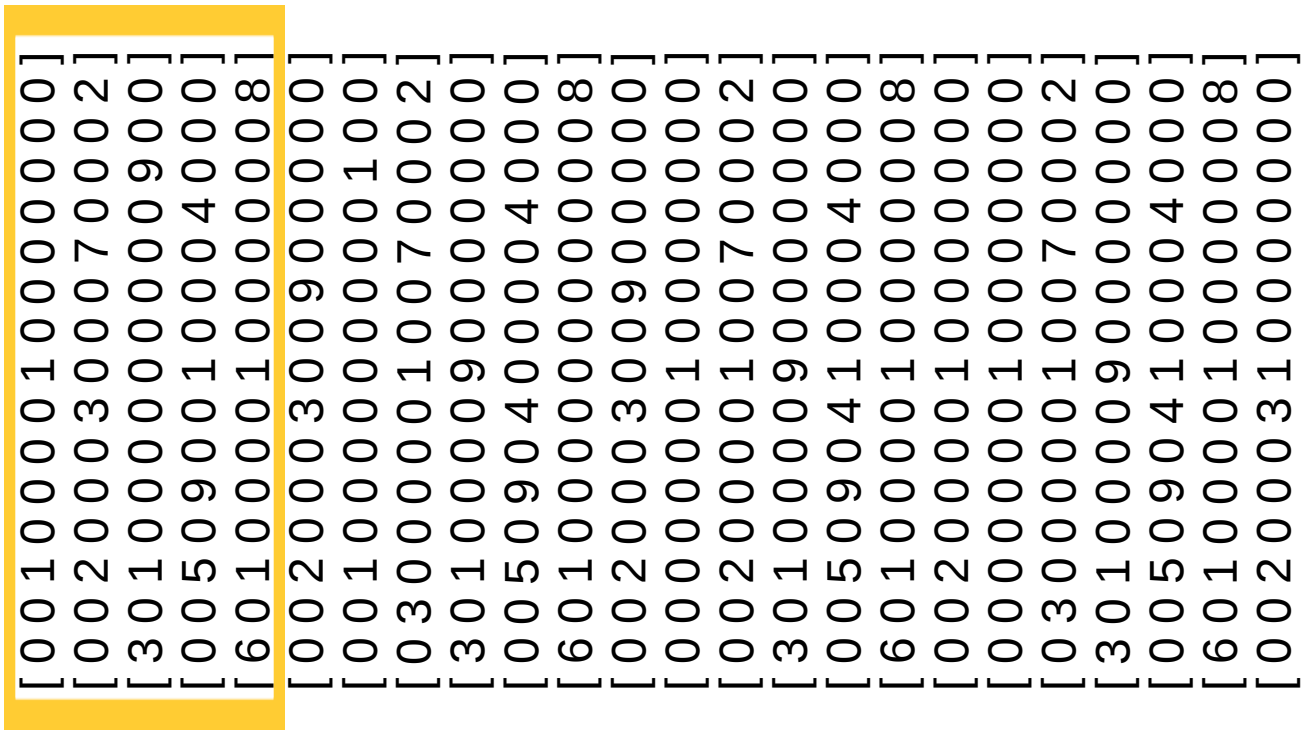
[0000010040000000010000700000001000000000]

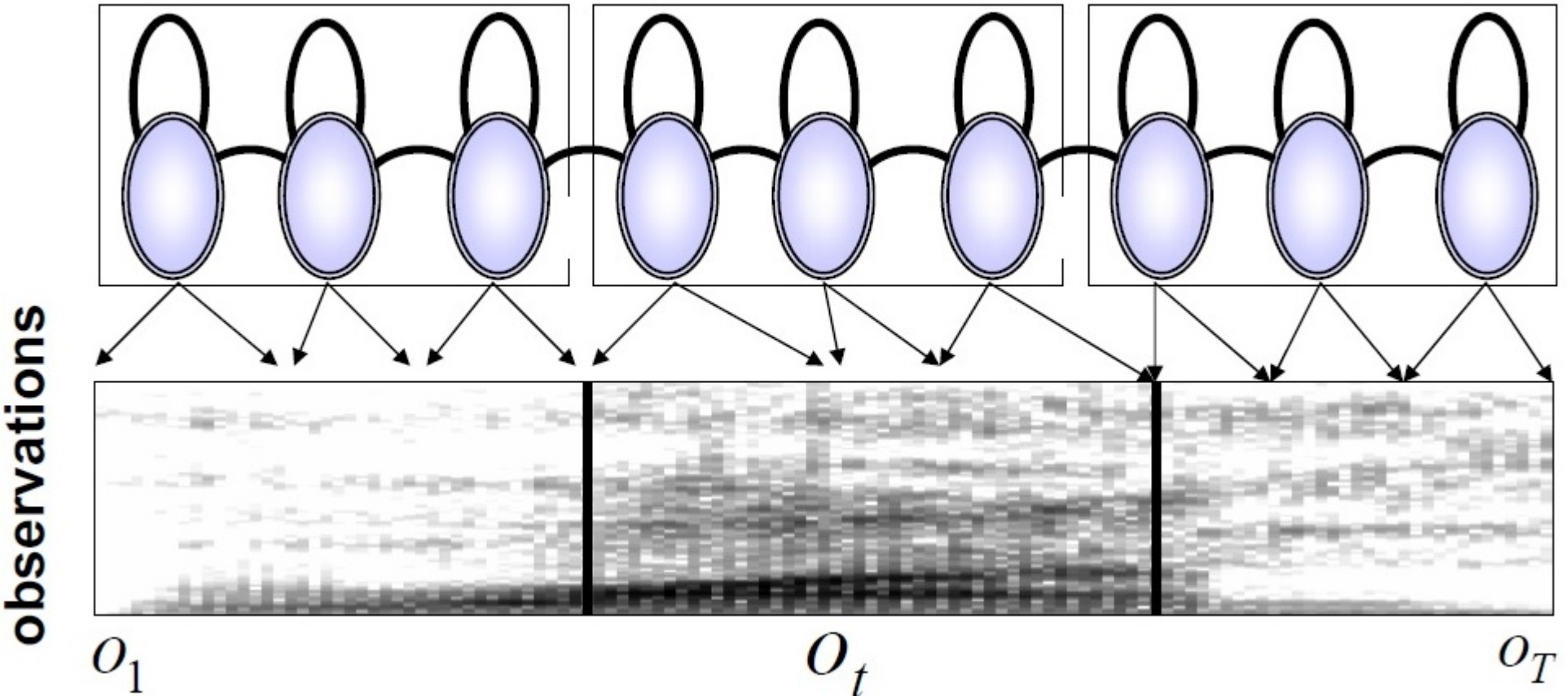
[0000010000500000010000000090001000000000]

.

.

HTK ALIGNMENT





<vsooda.github.io>

WORD →

<token text="илнй" 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">

<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"/>

<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"/>

<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"/>

<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"/>

</token>

WORD → <token text="илнё" 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 →
<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"/>

PHONEME →
<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"/>

PHONEME →
<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"/>

PHONEME →
<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"/>

</token>

WORD → <token text="илнё" 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">

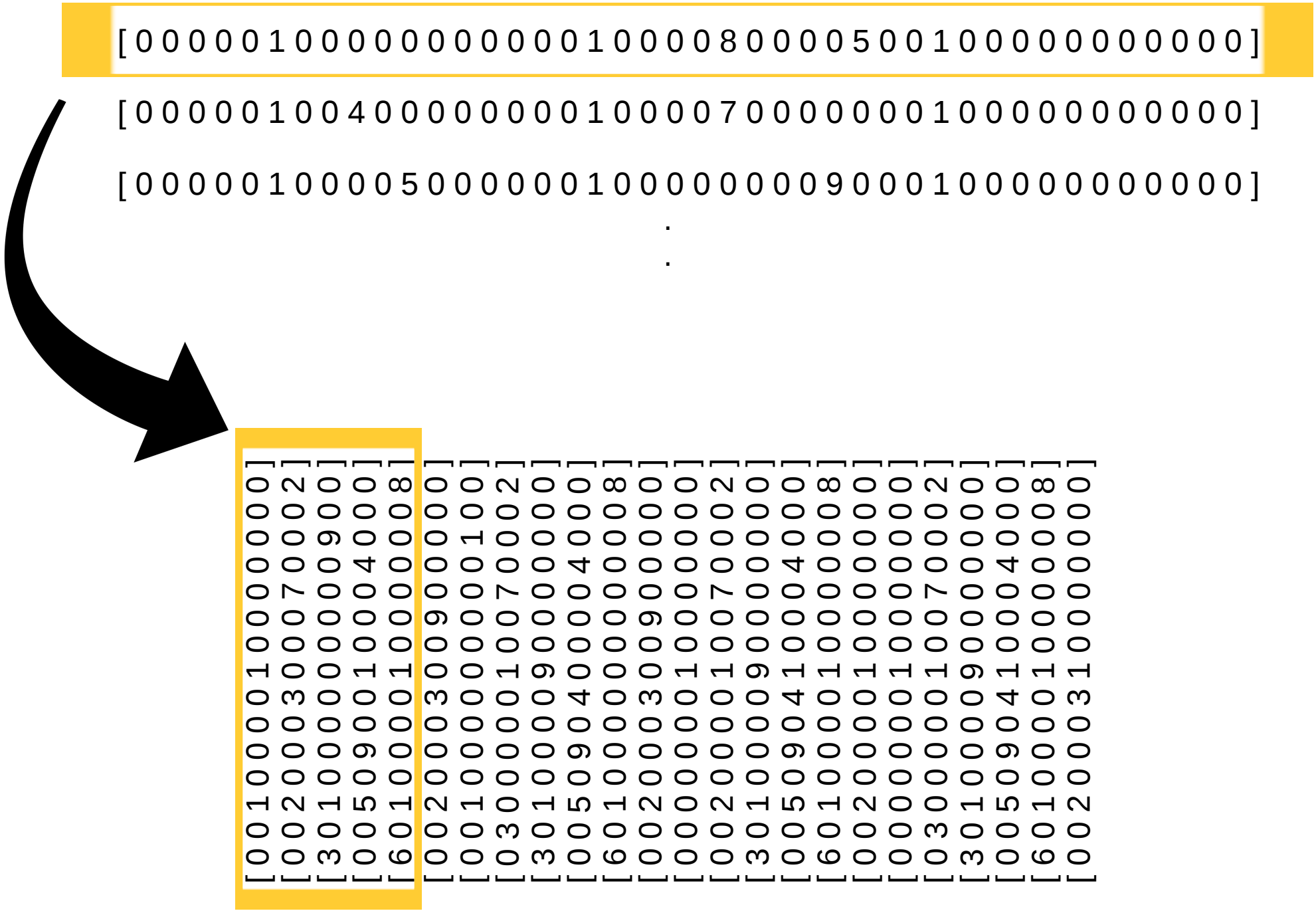
PHONEME →
<state start="2455" end="2460"/>
<state start="2460" end="2475"/>
<state start="2475" end="2500"/> ← TIME STAMPS
<state start="2500" end="2505"/>
<state start="2505" end="2510"/>

PHONEME →
<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"/>

PHONEME →
<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"/>

PHONEME →
<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"/>

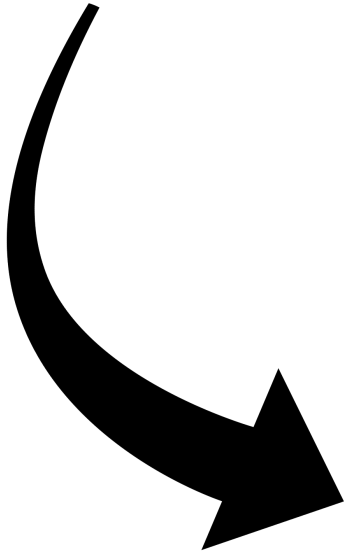
</token>


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

[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0]

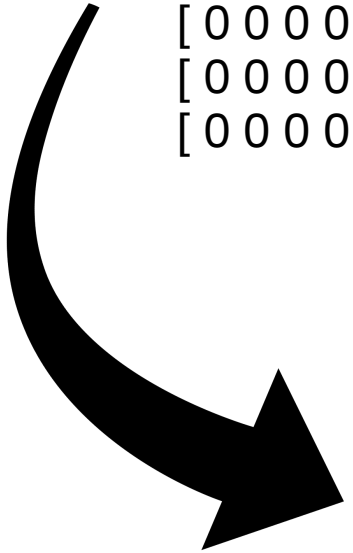
.

.



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

[0000010000000000001000080000500100000000000]
[0000010000000000001000080000500100000000000]
[0000010000000000001000080000500100000000000]
[0000010000000000001000080000500100000000000]
[0000010000000000001000080000500100000000000]

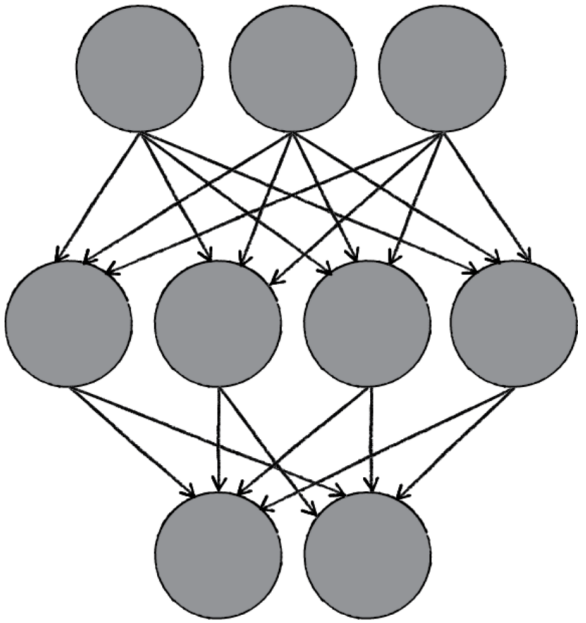


[00100000100000000]
[00200030000000002]
[30100000000900003]
[00509000100000000]
[60100000100000008]

```
[ 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 ]
[ 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 ]
[ 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 ]
[ 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 ]
[ 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 ]
```

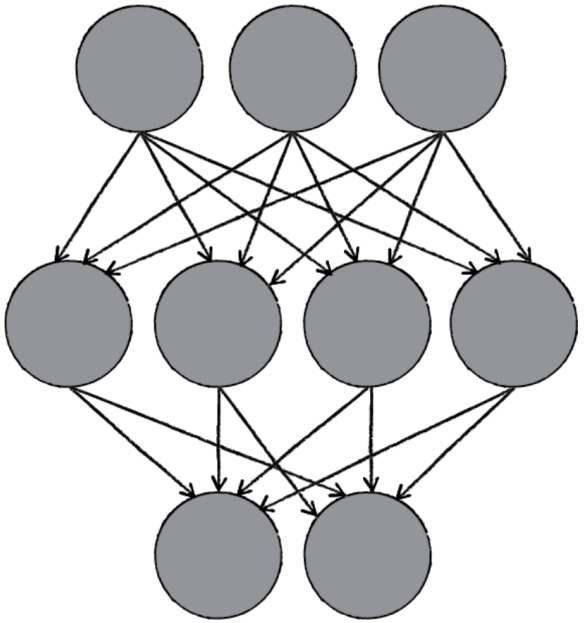
```
[ 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 ]
[ 0 0 2 0 0 0 3 0 0 0 7 0 0 0 2 ]
[ 3 0 1 0 0 0 0 0 0 0 0 0 9 0 0 ]
[ 0 0 5 0 9 0 0 1 0 0 0 4 0 0 0 ]
[ 6 0 1 0 0 0 0 1 0 0 0 0 0 0 8 ]
```

[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0 0]
[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0 0]
[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0 0]
[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0 0]
[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0 0]



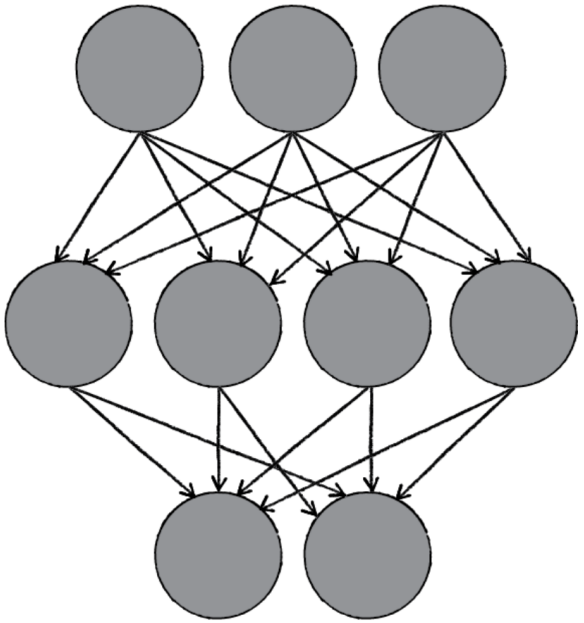
[0 0 1 0 0 0 0 1 0 0 0 0 0 0 0]
[0 0 2 0 0 0 3 0 0 0 7 0 0 0 2]
[3 0 1 0 0 0 0 0 0 0 0 0 9 0 0]
[0 0 5 0 9 0 0 1 0 0 0 4 0 0 0]
[6 0 1 0 0 0 0 1 0 0 0 0 0 0 8]

[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0 0]
[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0 0]
[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0 0]
[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0 0]
[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0 0]



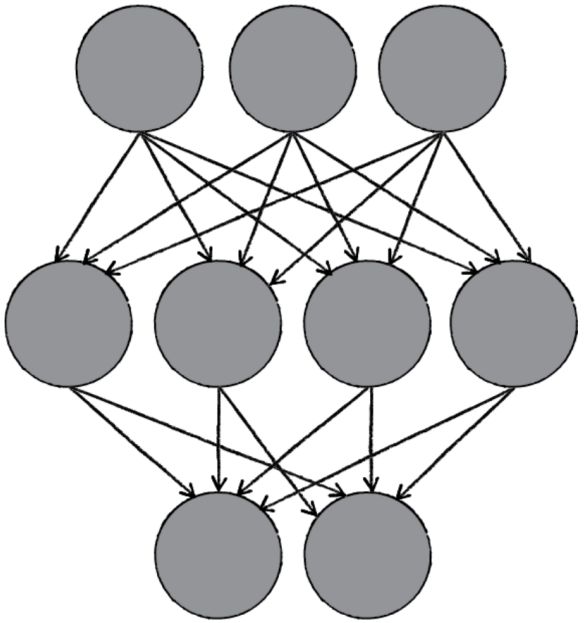
[0 0 1 0 0 0 0 1 0 0 0 0 0 0 0]
[0 0 2 0 0 0 3 0 0 0 7 0 0 0 2]
[3 0 1 0 0 0 0 0 0 0 0 0 9 0 0]
[0 0 5 0 9 0 0 1 0 0 0 4 0 0 0]
[6 0 1 0 0 0 0 1 0 0 0 0 0 0 8]

[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0 0]
[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0 0]
[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0 0]
[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0 0]
[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0 0]



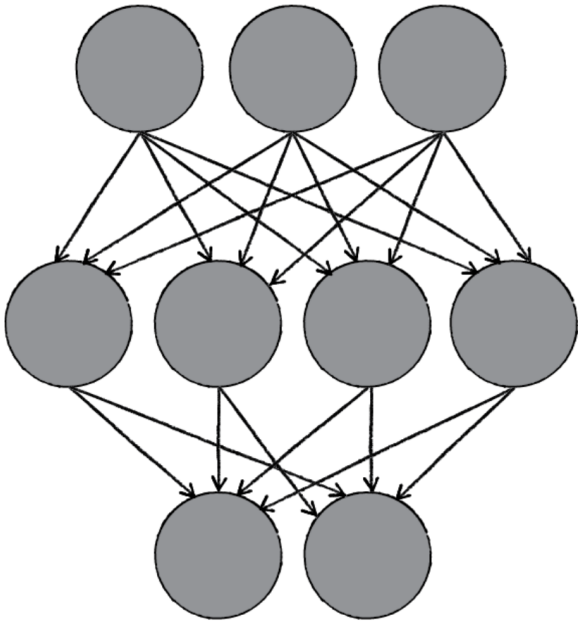
[0 0 1 0 0 0 0 1 0 0 0 0 0 0 0]
[0 0 2 0 0 0 3 0 0 0 7 0 0 0 2]
[3 0 1 0 0 0 0 0 0 0 0 0 9 0 0]
[0 0 5 0 9 0 0 1 0 0 0 4 0 0 0]
[6 0 1 0 0 0 0 1 0 0 0 0 0 0 8]

[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0 0]
[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0 0]
[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0 0]
[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0 0]
[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0 0]



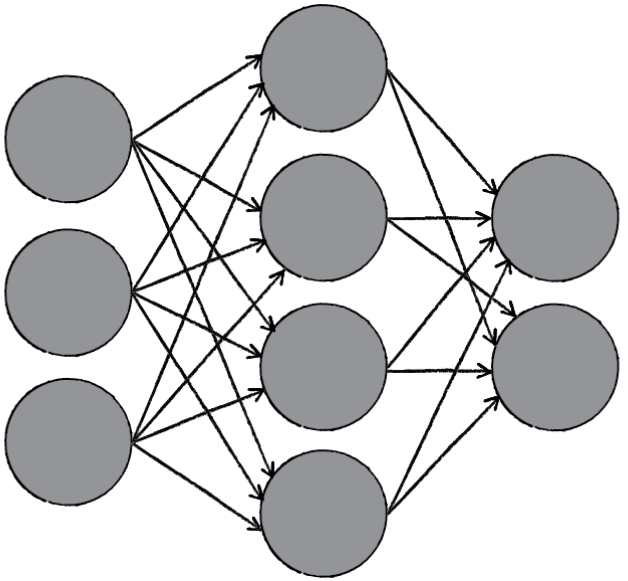
[0 0 1 0 0 0 0 1 0 0 0 0 0 0 0]
[0 0 2 0 0 0 3 0 0 0 7 0 0 0 2]
[3 0 1 0 0 0 0 0 0 0 0 0 9 0 0]
[0 0 5 0 9 0 0 1 0 0 0 4 0 0 0]
[6 0 1 0 0 0 0 1 0 0 0 0 0 0 8]

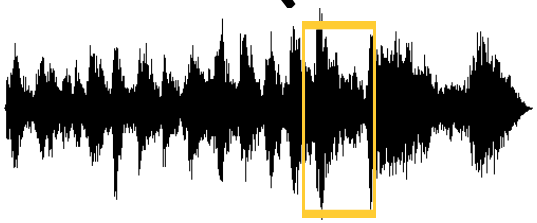
[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0 0]
[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0 0]
[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0 0]
[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0 0]
[0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 8 0 0 0 0 5 0 0 1 0 0 0 0 0 0 0 0 0 0 0]



[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 0 0 9 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]

Moscow always has
traffic jams.





References / Resources

- CSTR slides: speech.zone
- Ossian: github.com/CSTR-Edinburgh/Ossian
- Merlin: github.com/CSTR-Edinburgh