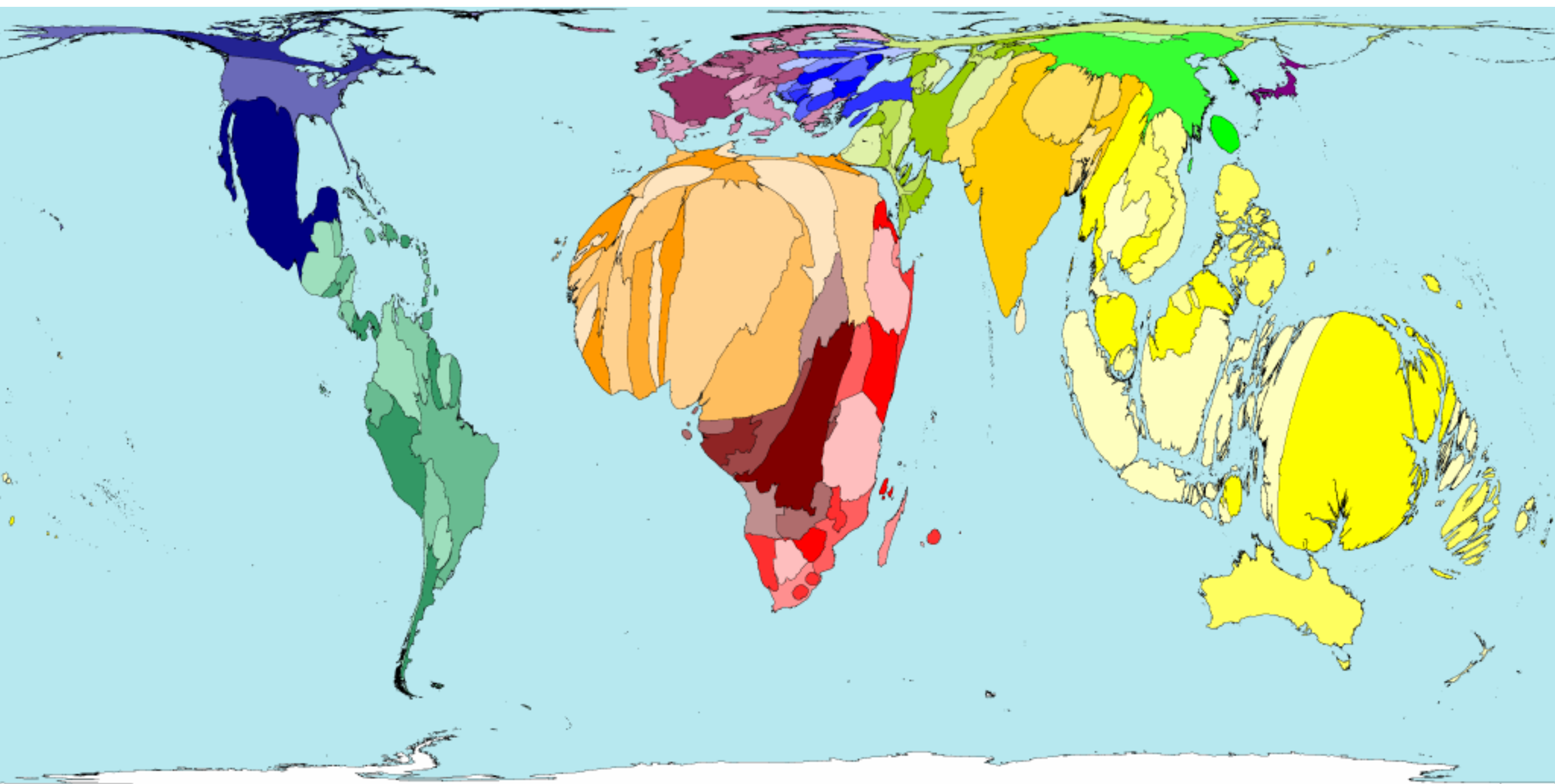


A case study on using speech-to-translation alignments for language documentation

Antonis Anastasopoulos, David Chiang

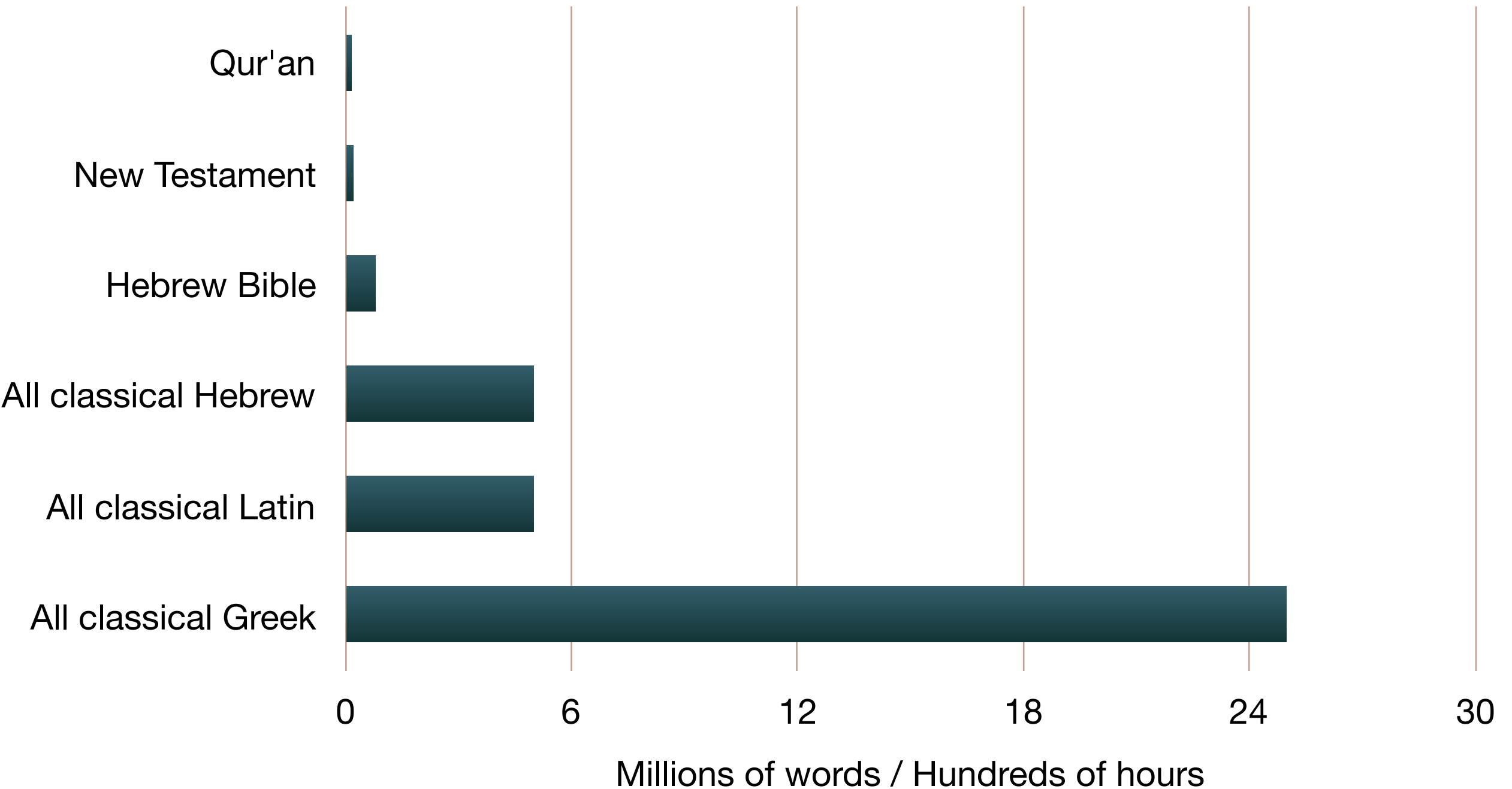




Goal

- Collect data now; analyze later
- The data must be:
 - Sufficient
 - Interpretable

How much data?



The linguistic graveyard



Making an audio Rosetta Stone

- Aikuma:
Android/web-
based app
- Push-to-talk,
push-to-
translate



Interpreting the audio Rosetta Stone



Interpreting the audio Rosetta Stone



Now their farm will not stay behind forever.

Interpreting the audio Rosetta Stone



Gila abur-u-n ferma hamišaluğ güğüna amuq'-da-č

Now their farm will not stay behind forever.

Interpreting the audio Rosetta Stone

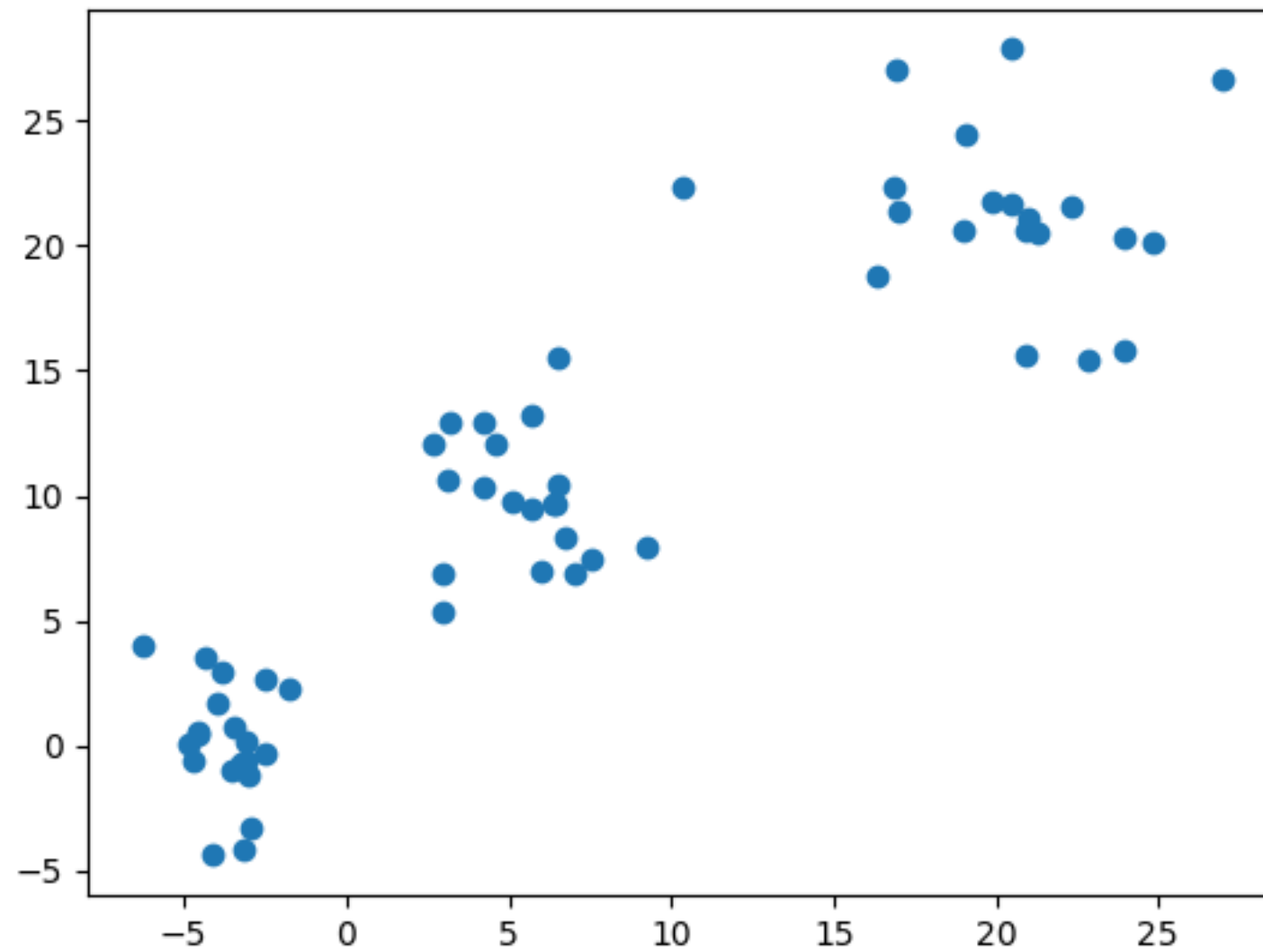


now their farm will not stay behind forever

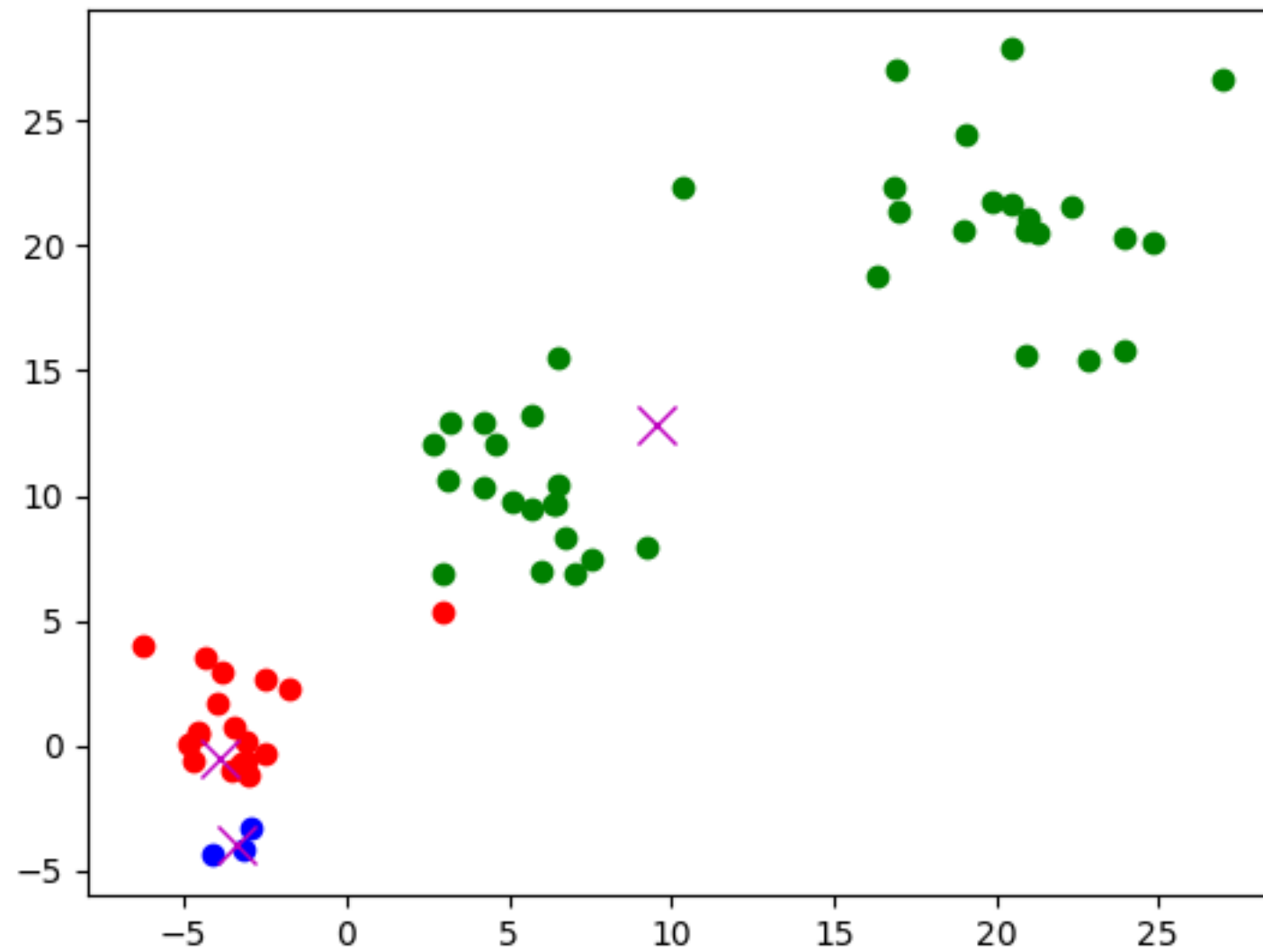
Background

K-means Clustering

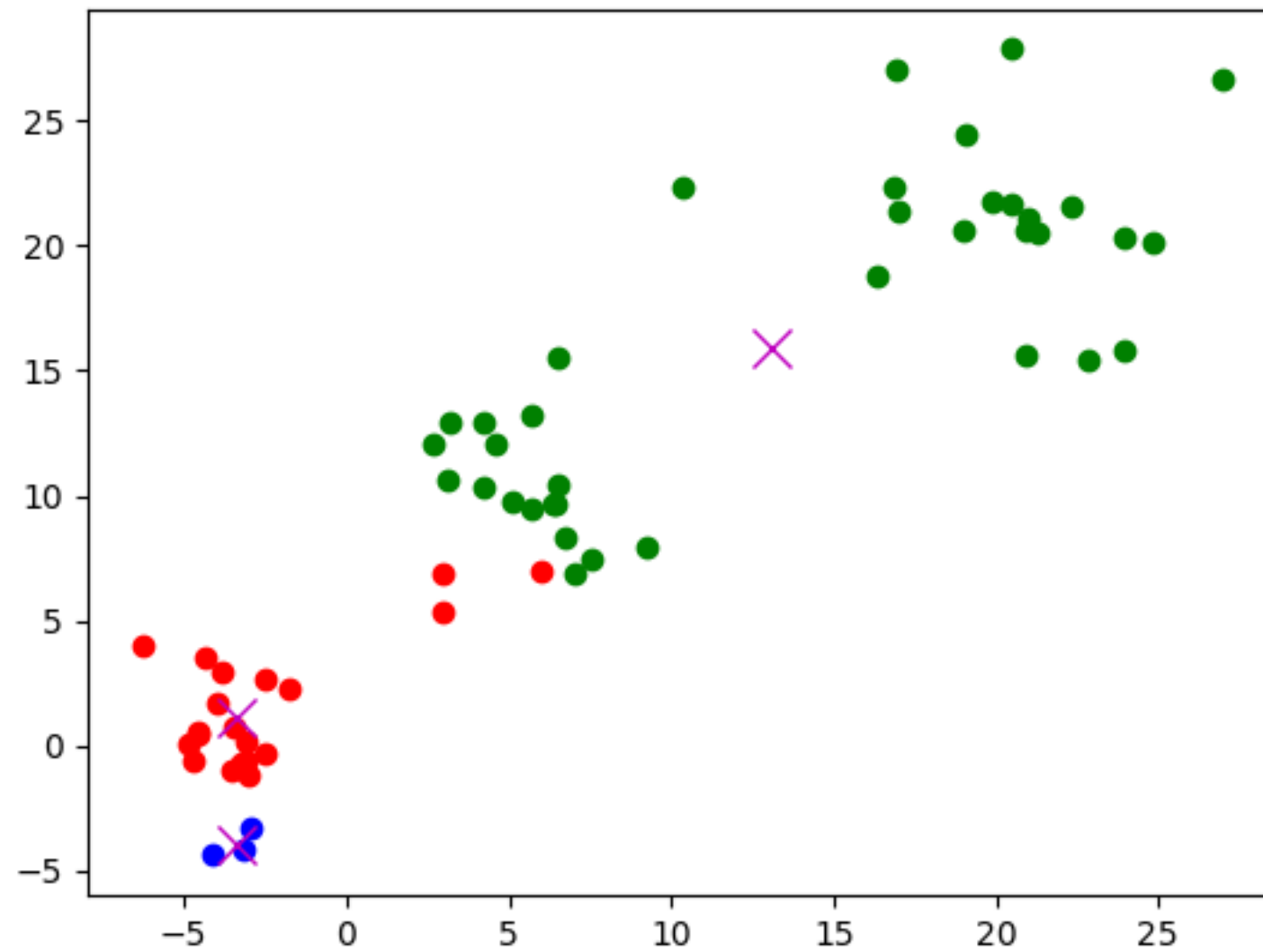
K-means Clustering



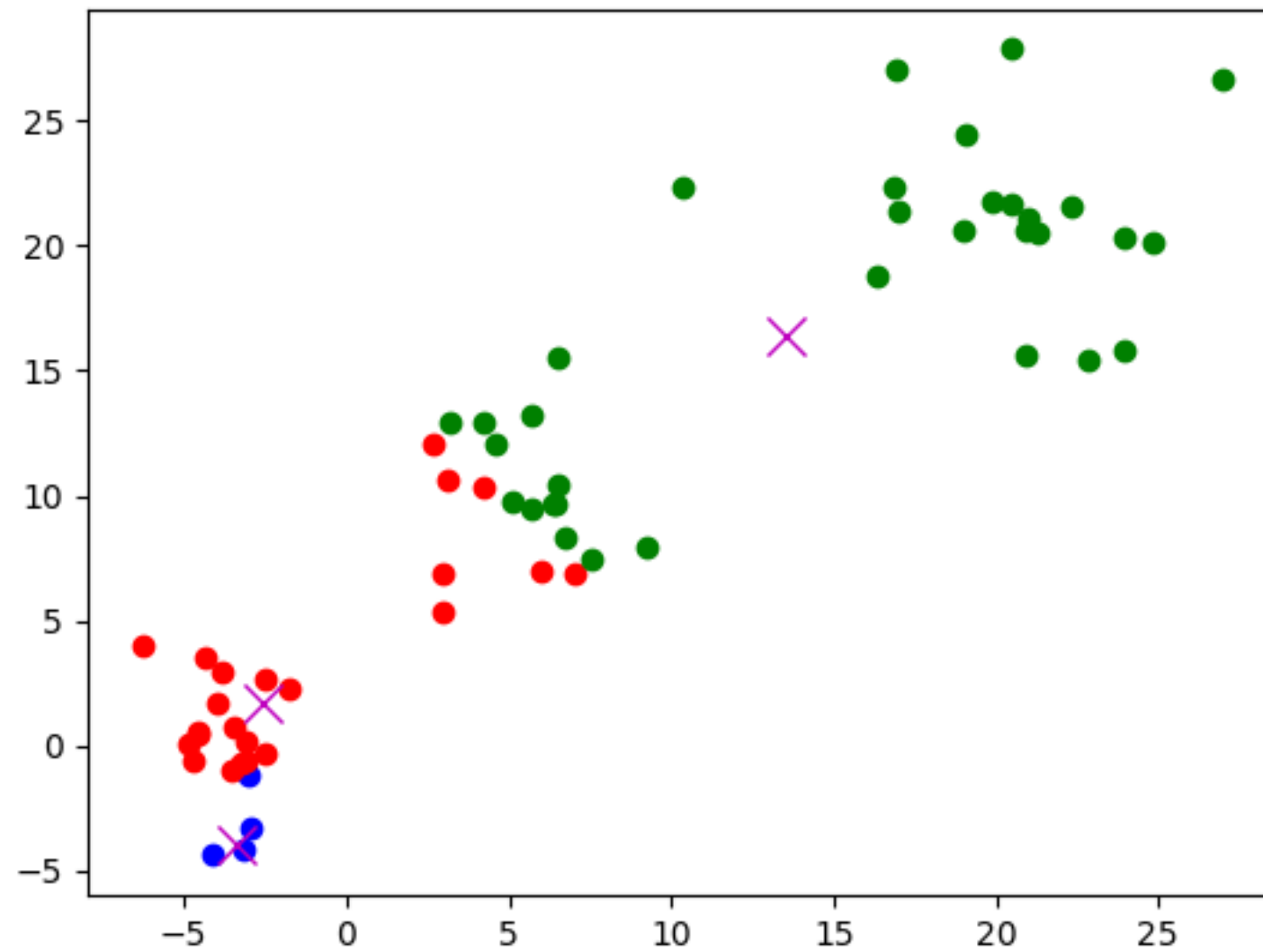
K-means Clustering



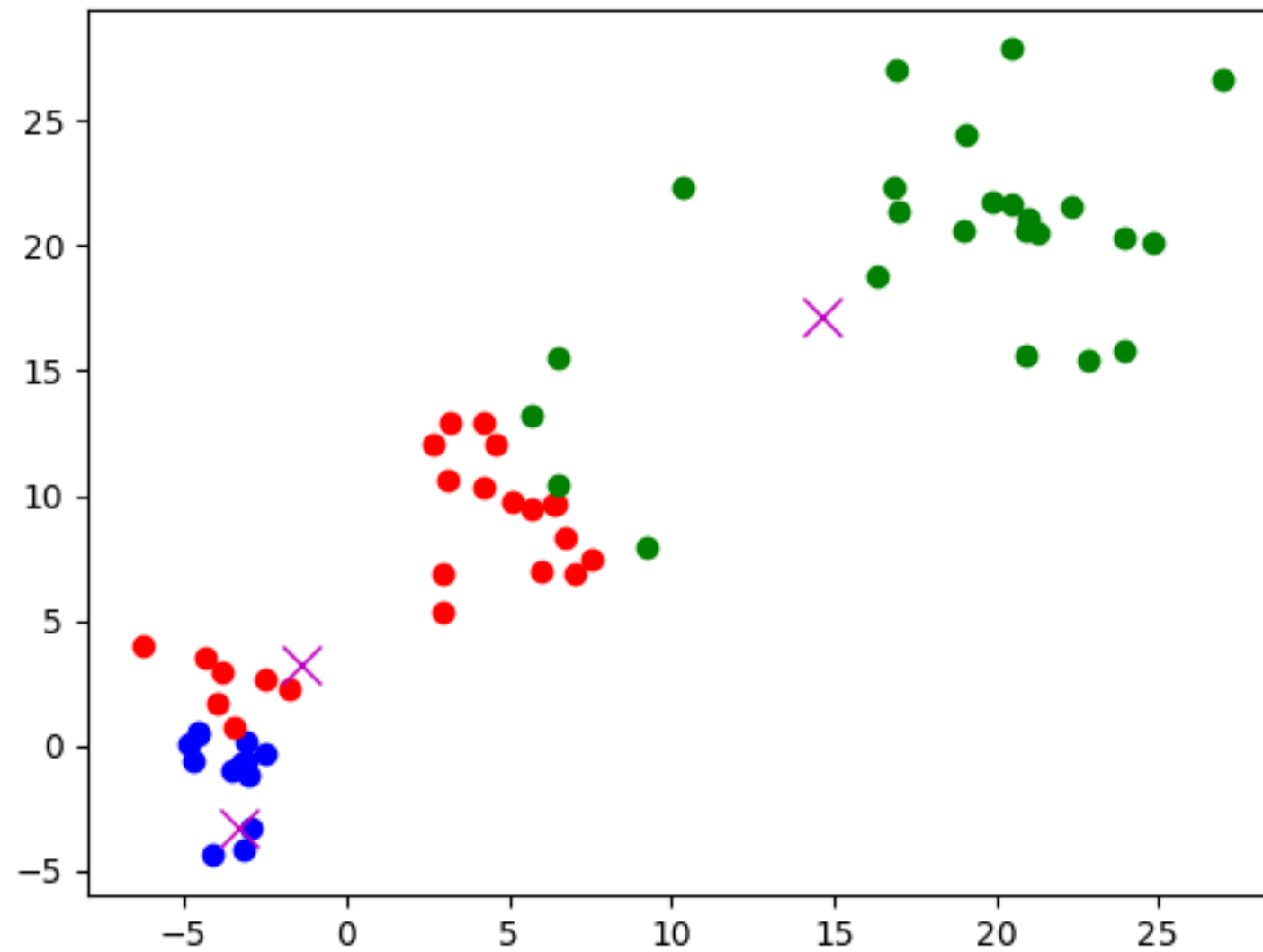
K-means Clustering



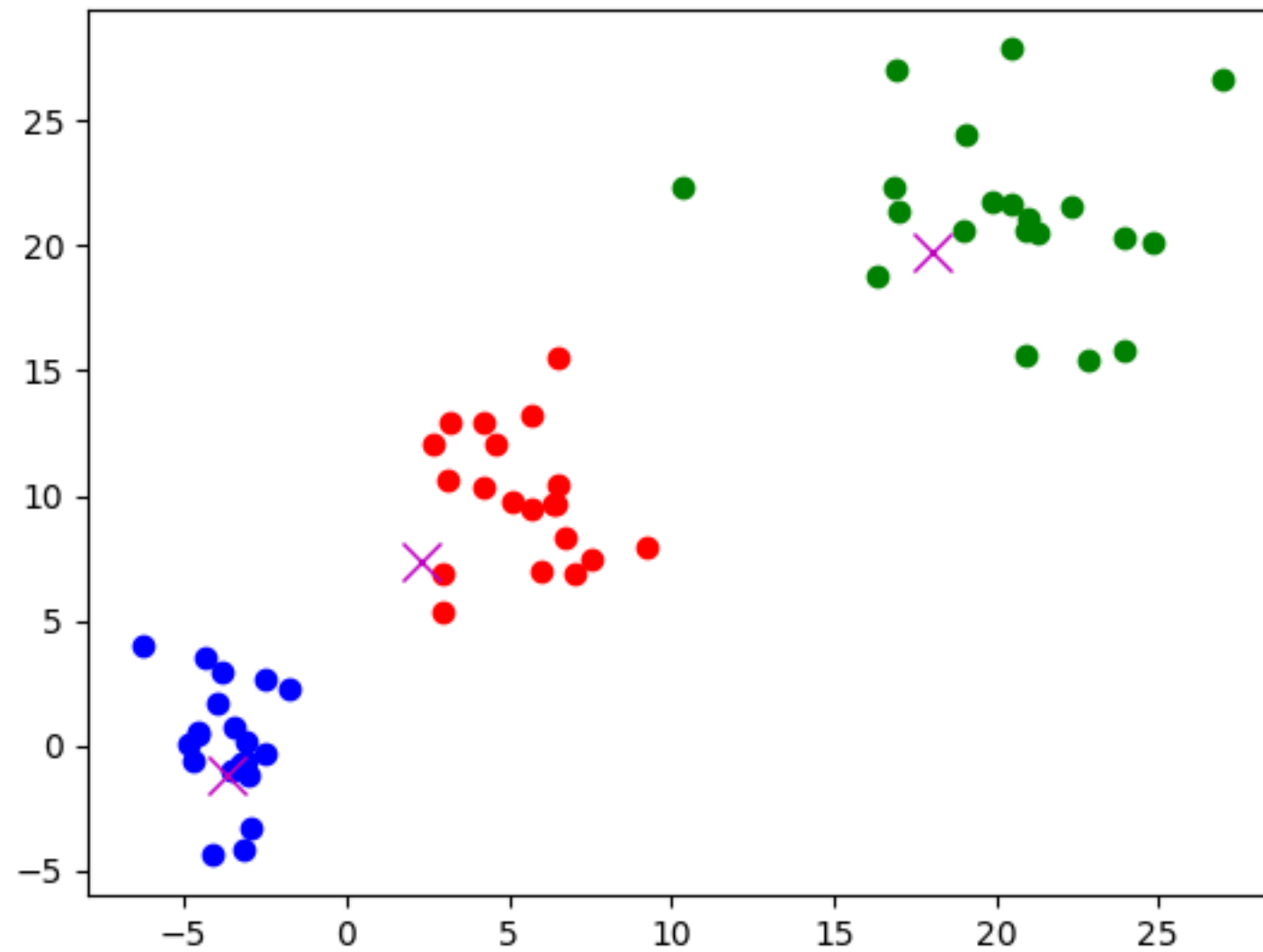
K-means Clustering



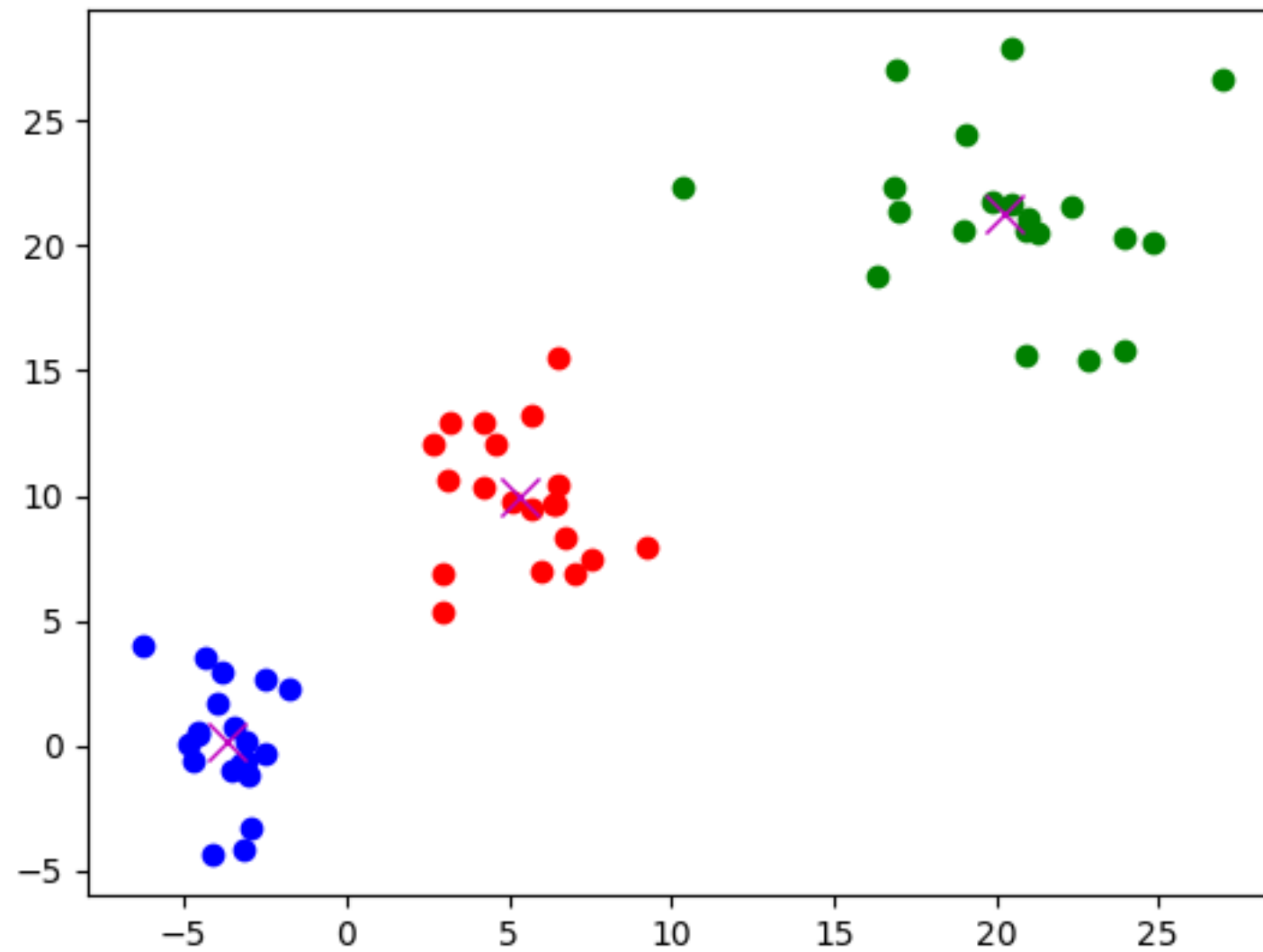
K-means Clustering



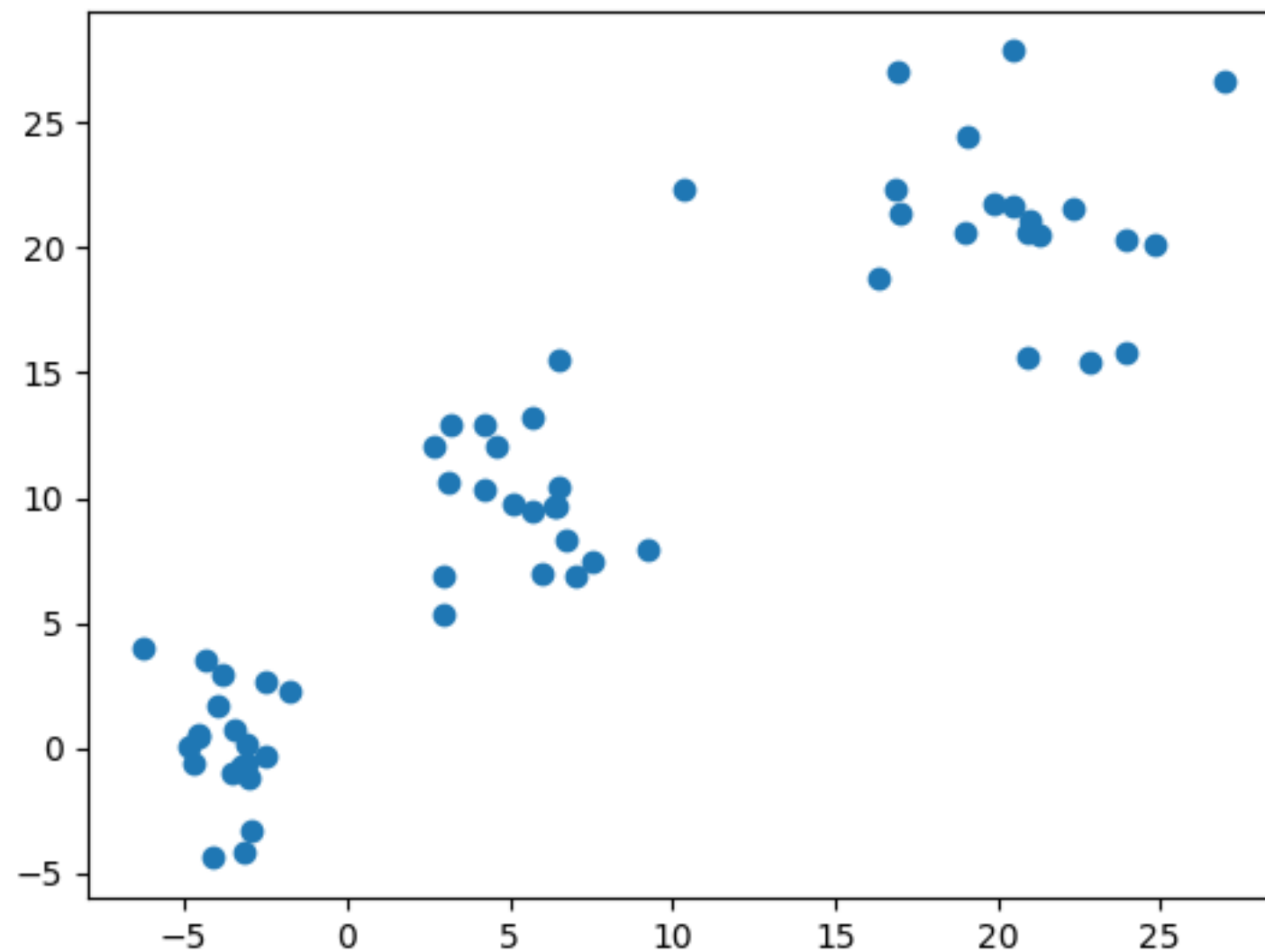
K-means Clustering



K-means Clustering

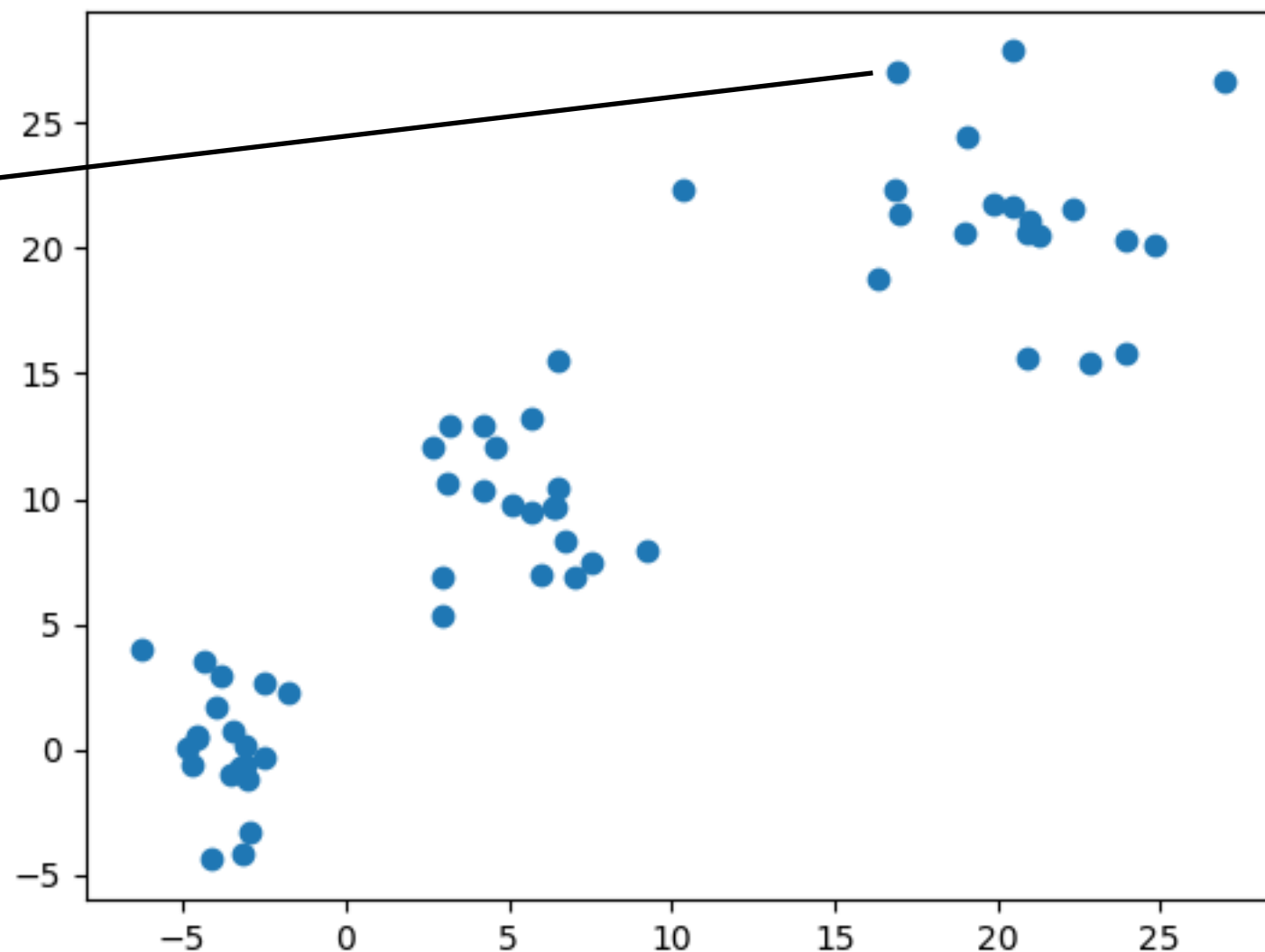
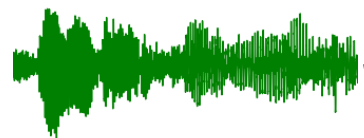


Aligning Speech to Translation

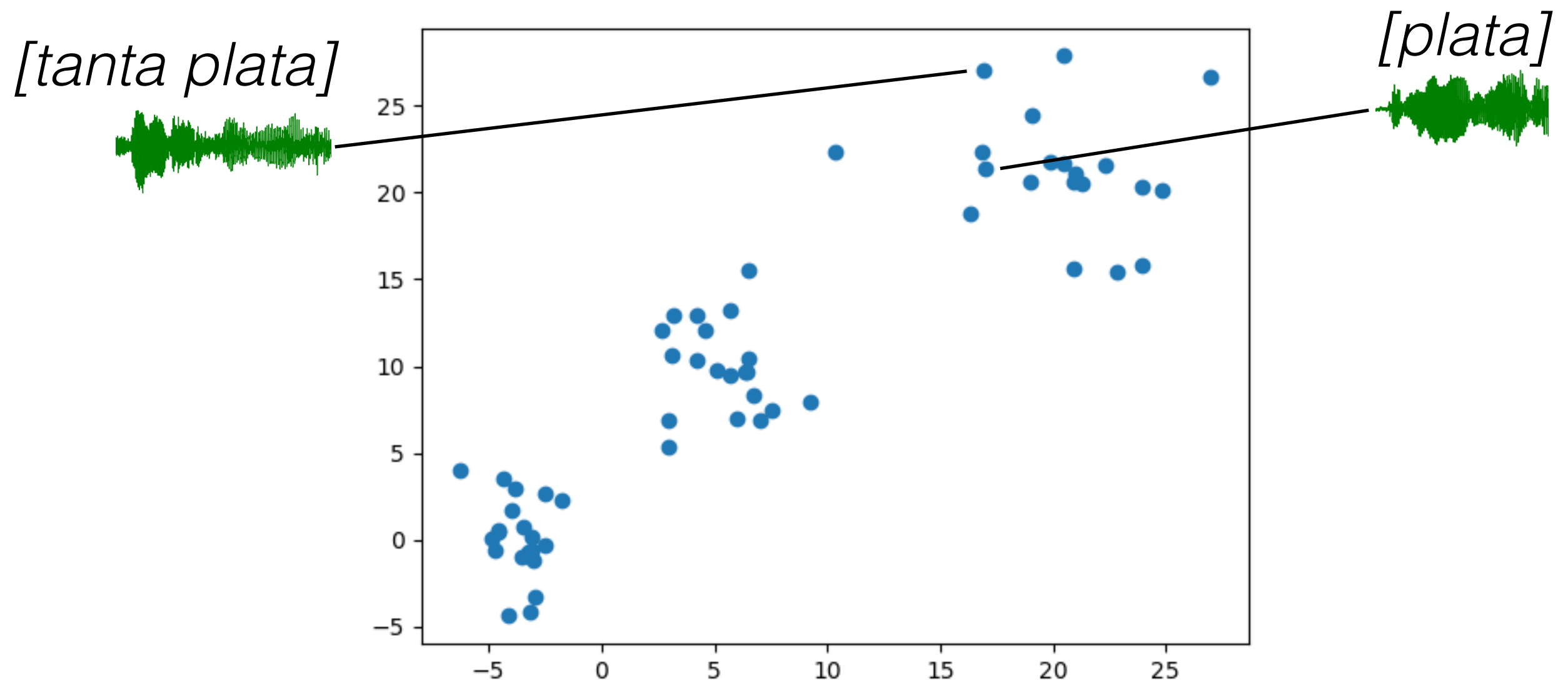


Aligning Speech to Translation

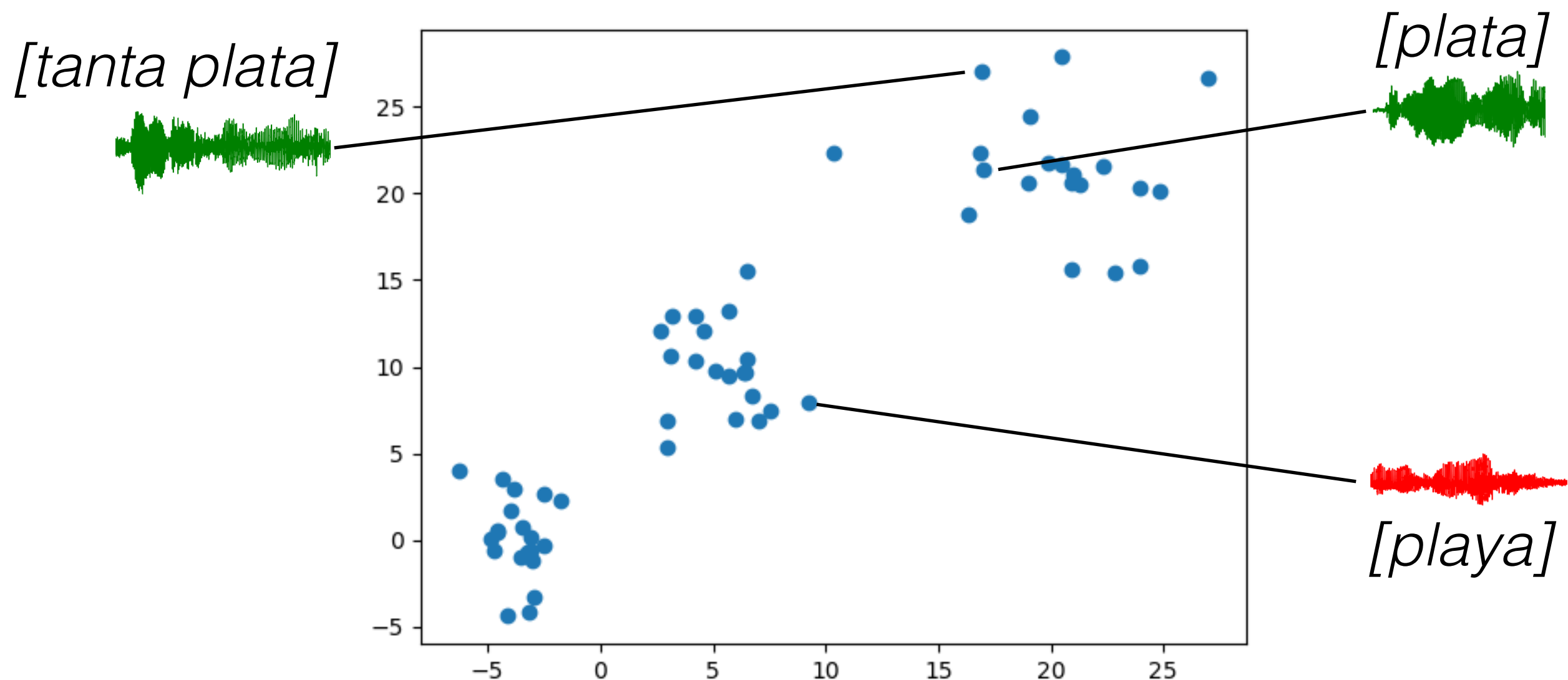
[tanta plata]



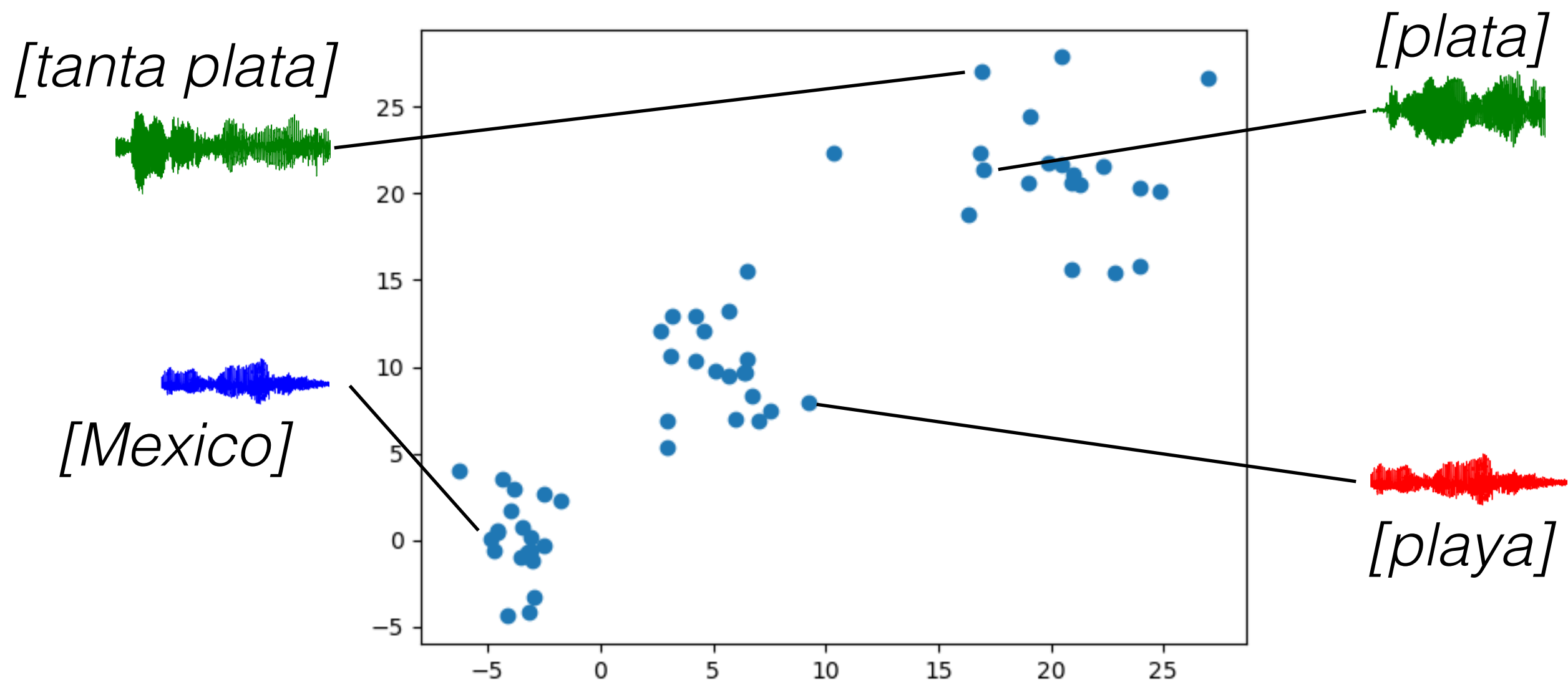
Aligning Speech to Translation



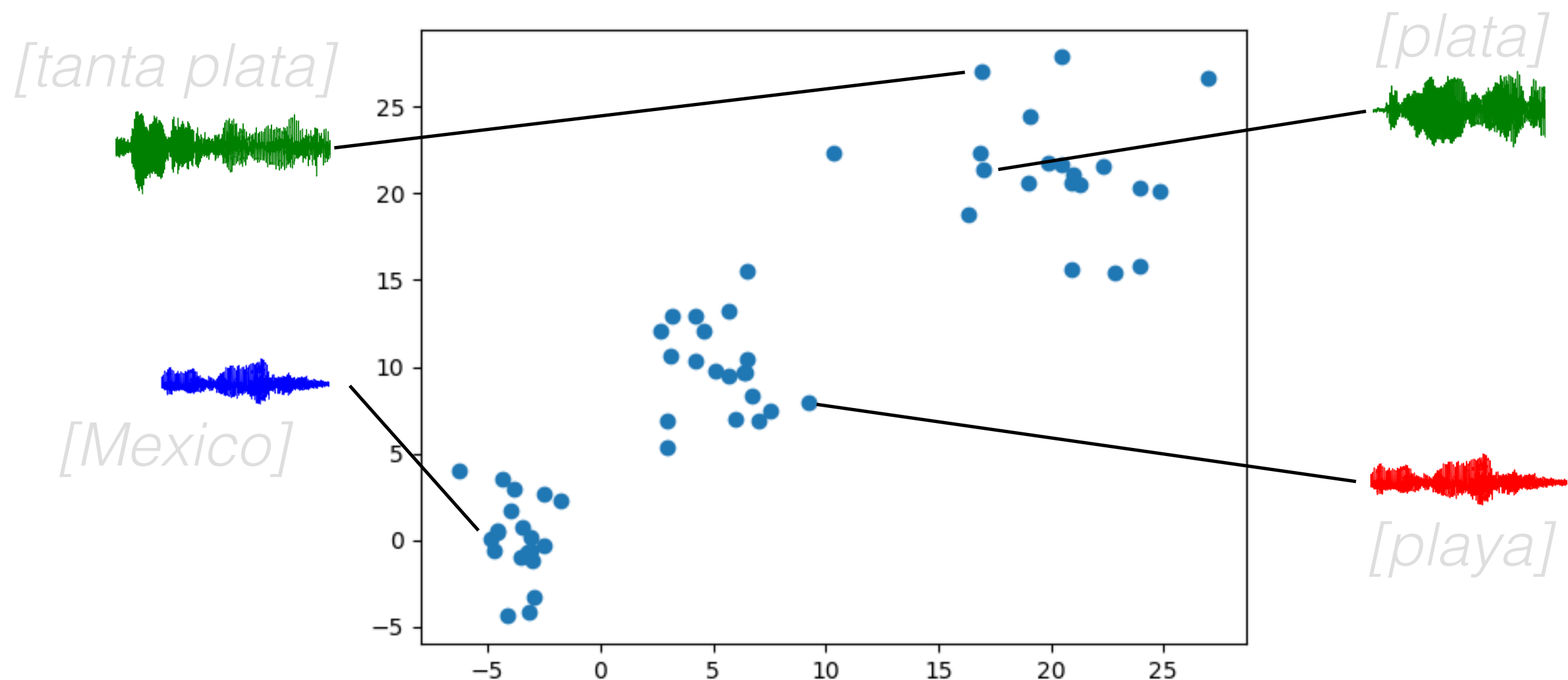
Aligning Speech to Translation



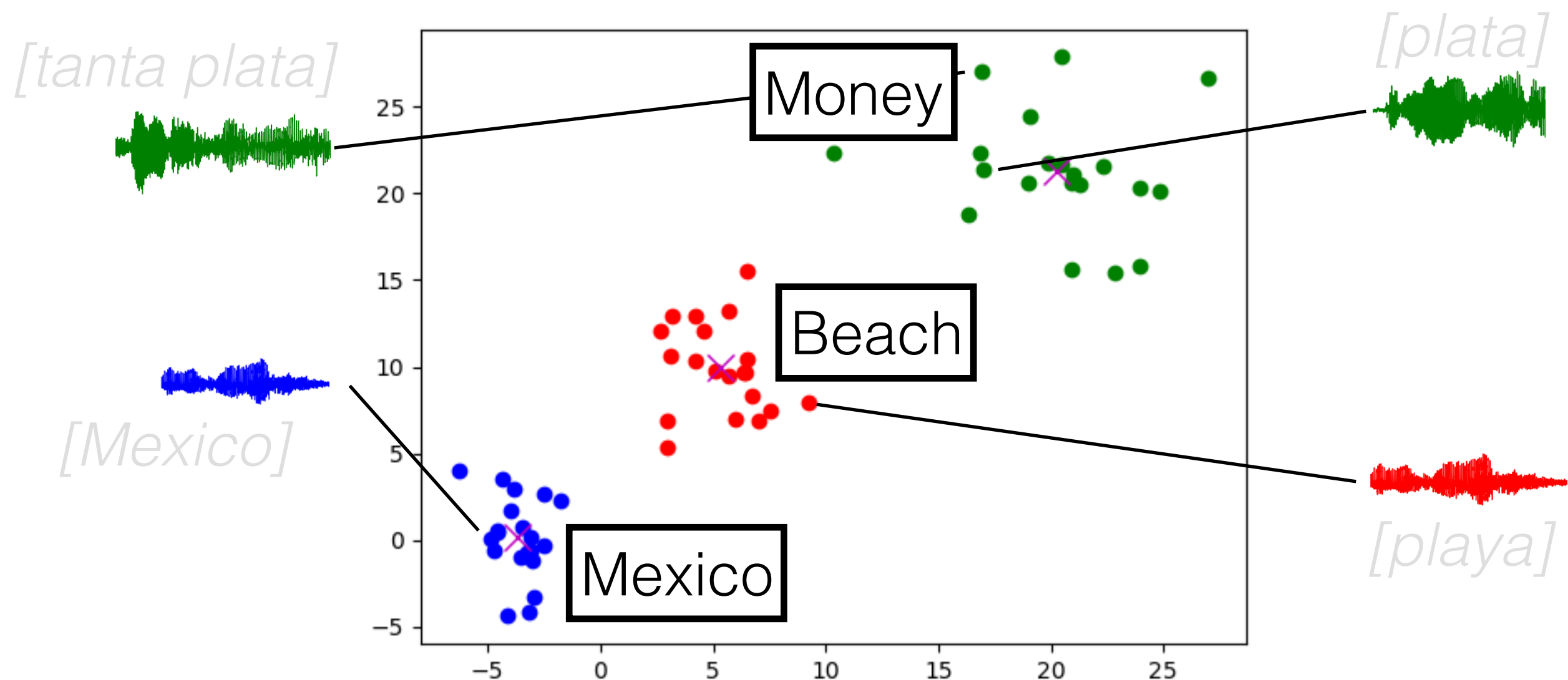
Aligning Speech to Translation



Aligning Speech to Translation

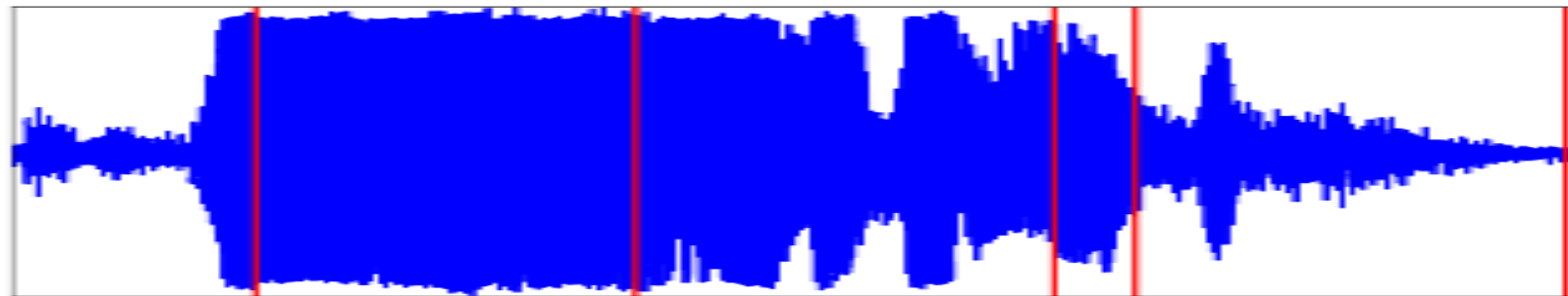


Aligning Speech to Translation



Example output

é Valeria meletá o' giornále

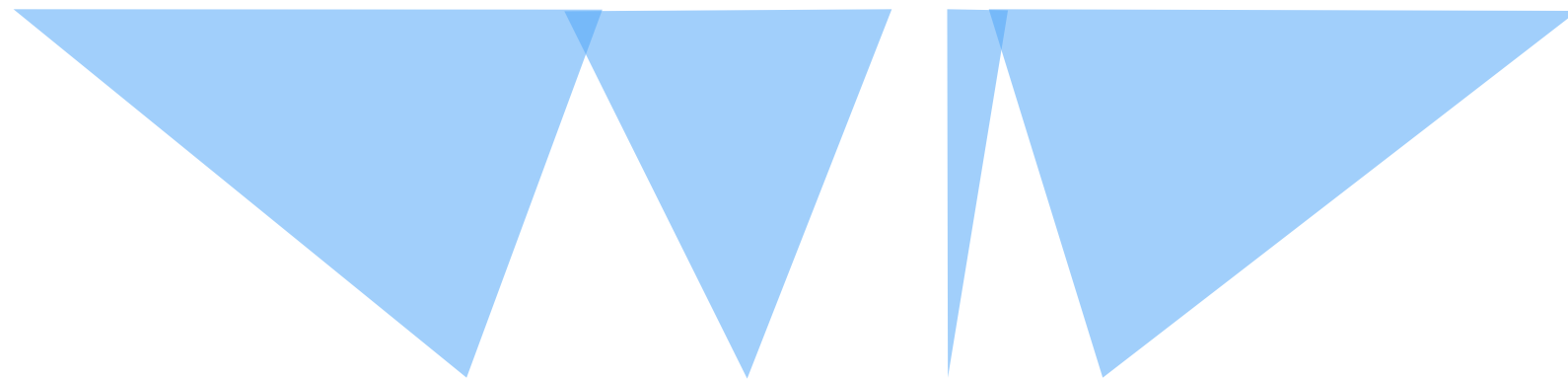
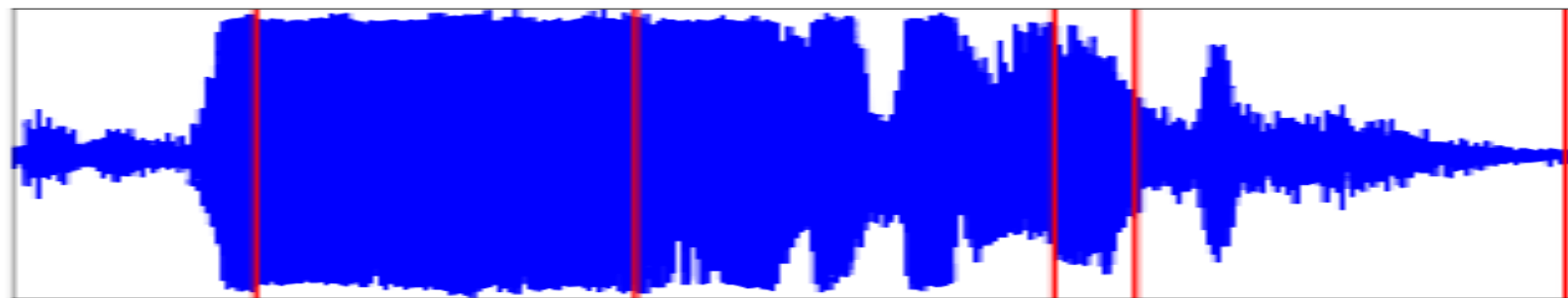


Valeria legge il giornale

[Valeria reads the newspaper]

Example output

é Valeria meletá o' giornále

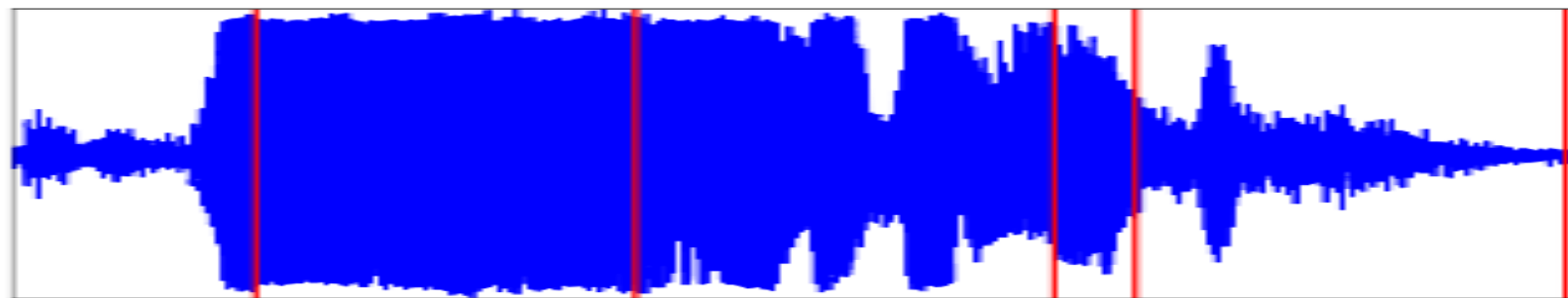


Valeria legge il giornale

[Valeria reads the newspaper]

Example output

é Valeria meletá o' giornále



Score: 0.82

Valeria legge il giornale

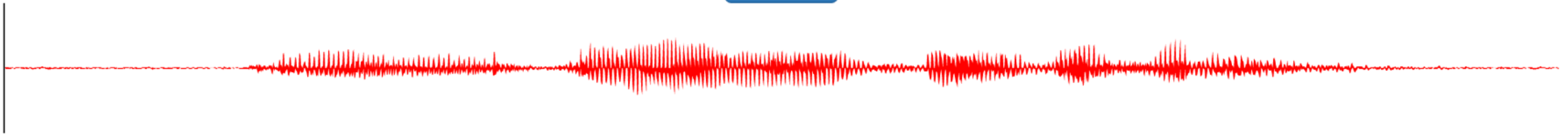
[Valeria reads the newspaper]

User Study

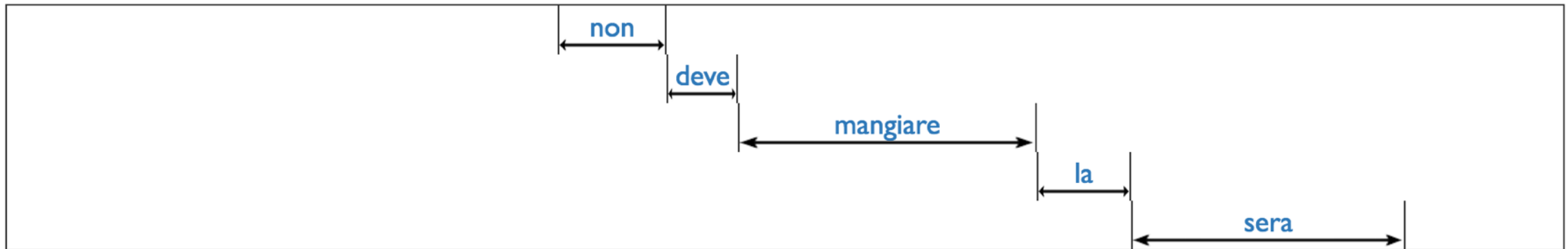
Griko



User Interface



Translation: non deve mangiare la sera

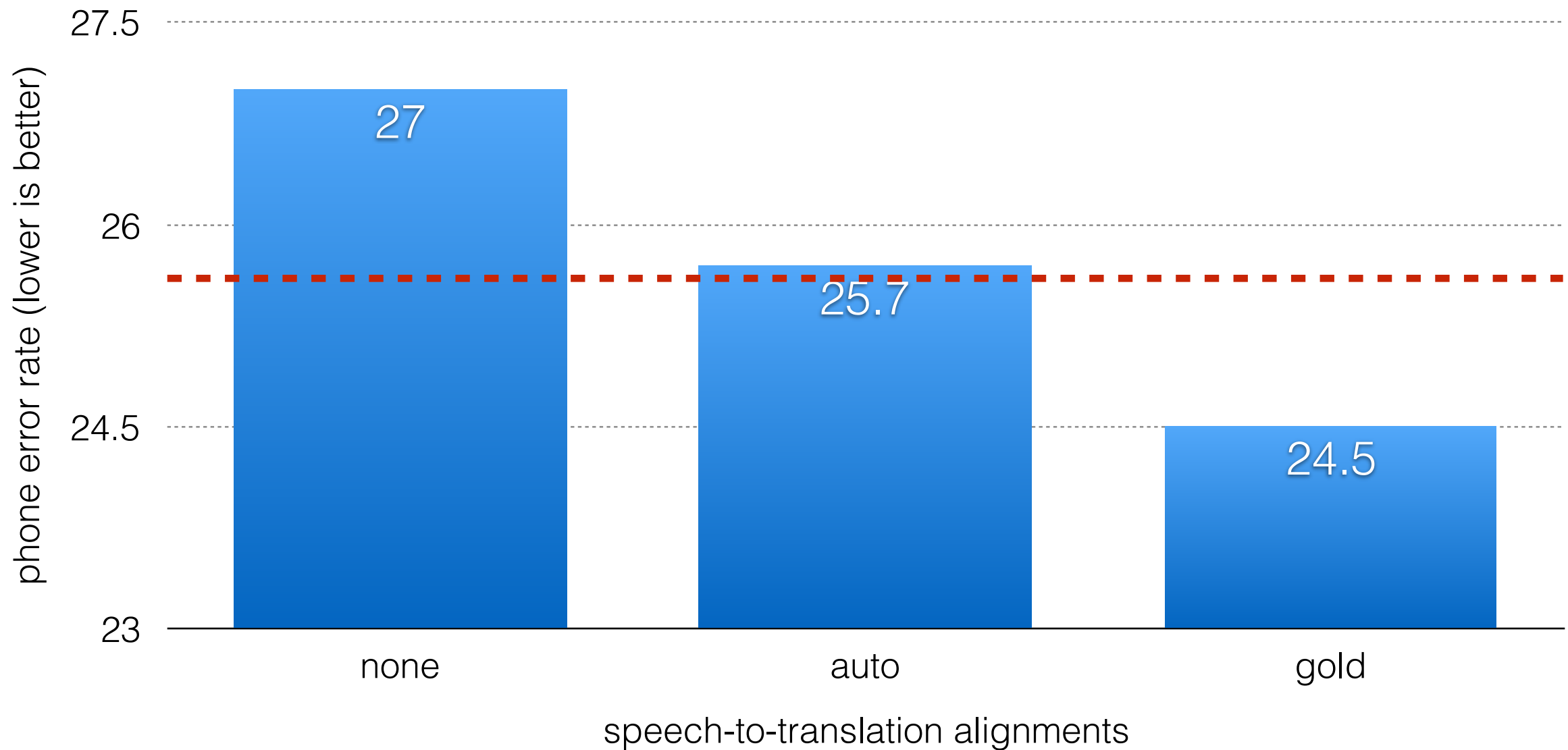


Transcription:

Save

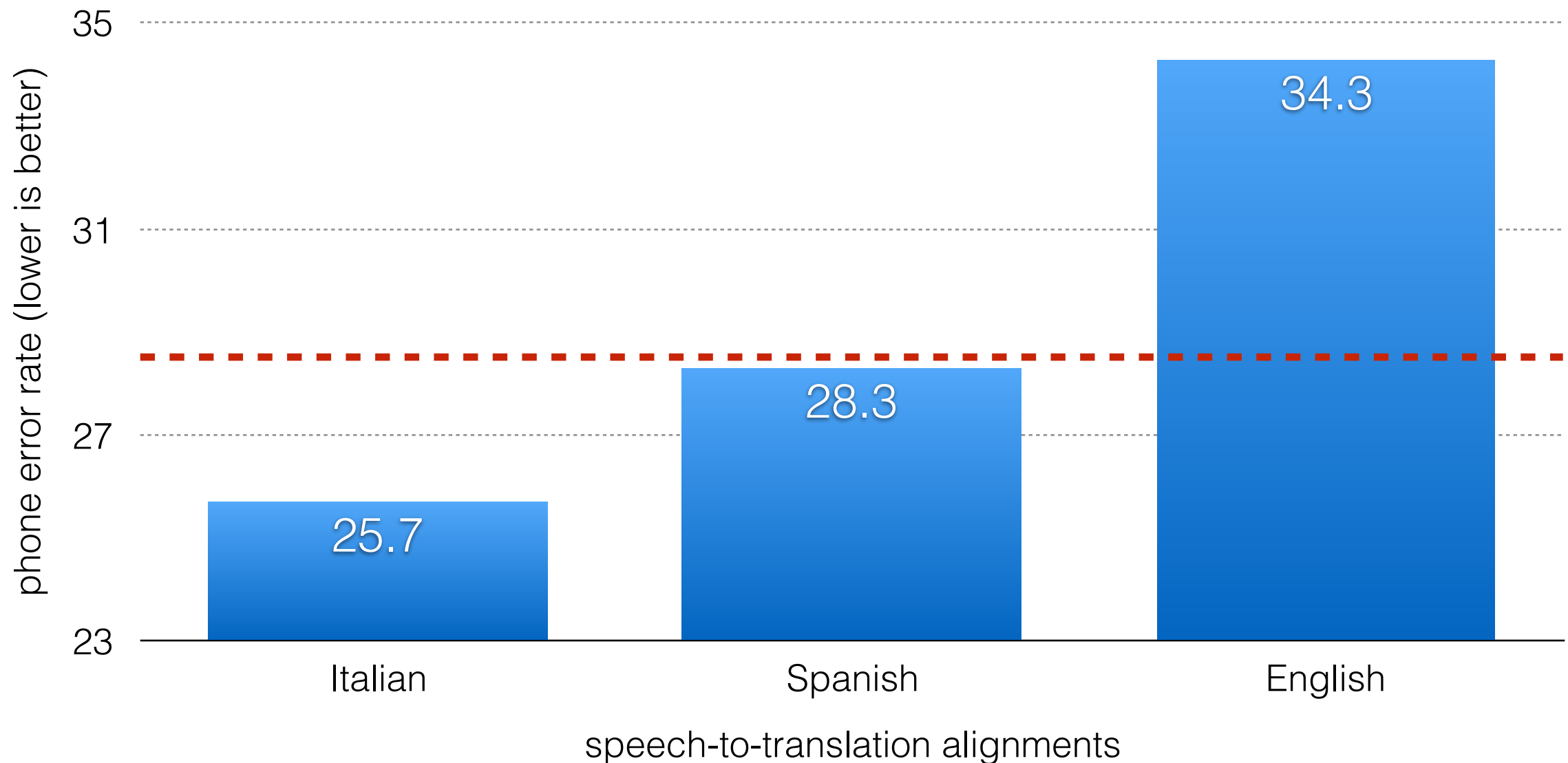
Results

error rates averaged
across 6 Italian-speaking participants



Results

error rates averaged across all participants
(6 Italian, 3 Spanish, 3 English)



Consensus transcription

user	transcription	distance
it1	<i>o ladro i so ndze mia buttu</i>	5
it2	<i>o ladro isodZenti dabol tu</i>	6
it3	<i>o ladro isodzeem biabiddu</i>	5
correct	<i>o ladro isodZe embi apo ttu</i>	

Can we do better?

Combine them!

String averaging

String averaging

oladroidndzemibuttu

oladroidZentidaboltu

oladroidzeembibiddu

String averaging

0	—	<i>o</i>	<i>l a d r o i s o n d z e m i a b u t t u</i>
1			
1			
...			
1			

o l a d r o i s o d Z e n t i d a b o l t u

o l a d r o i s o d z e e m b i a b i d d u

String averaging

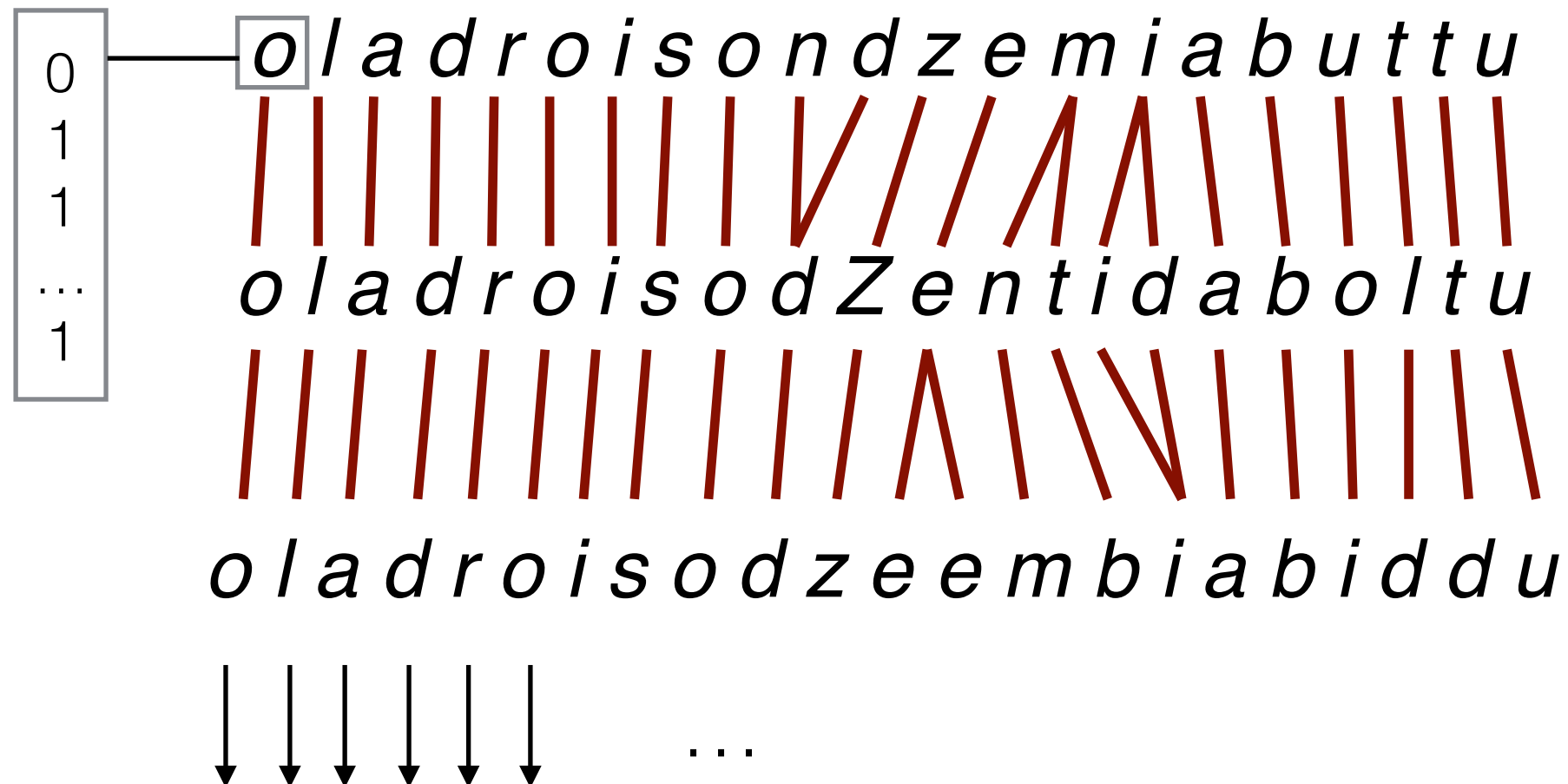


oladrisodzeembiabidu

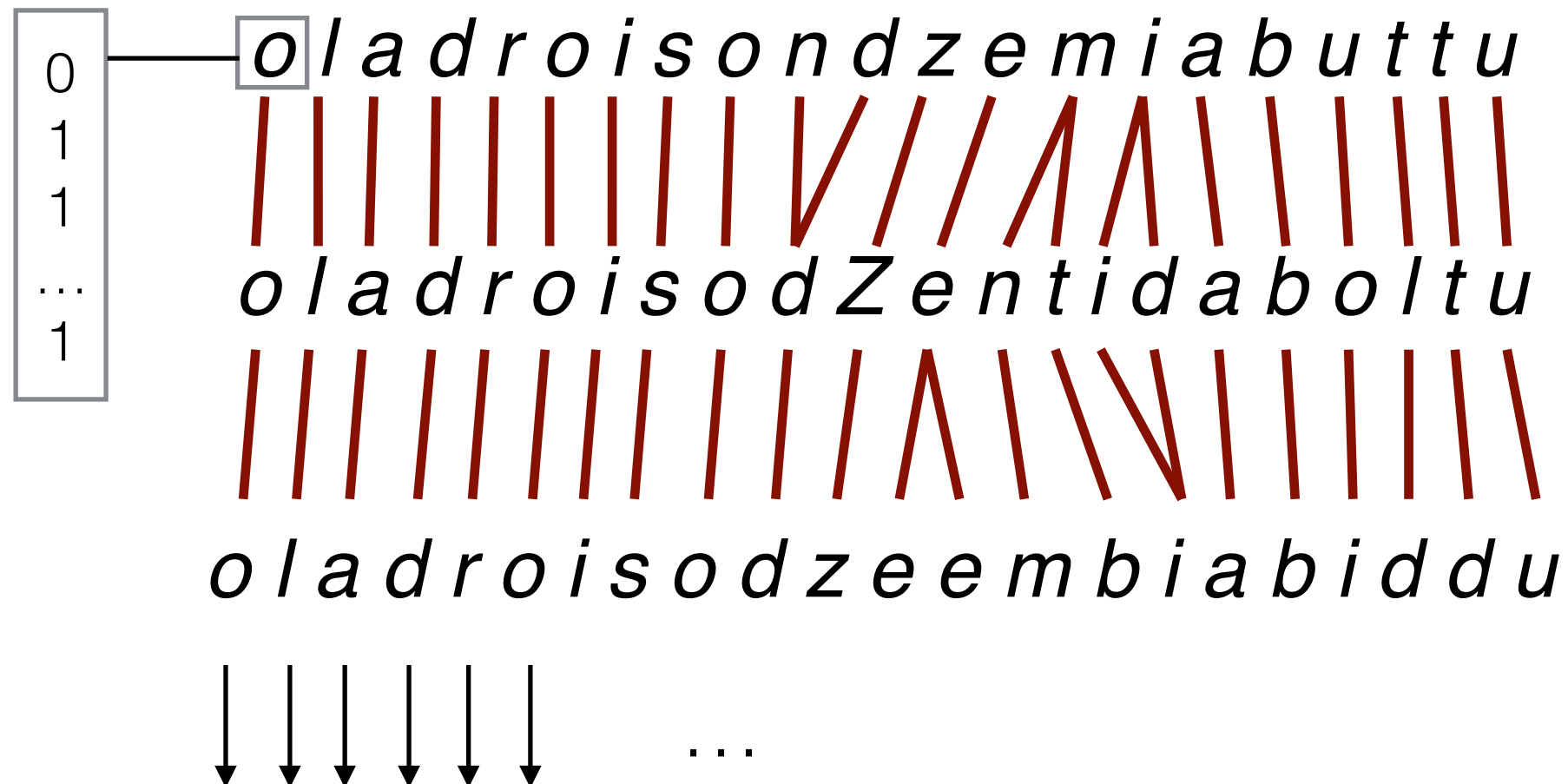
String averaging



String averaging

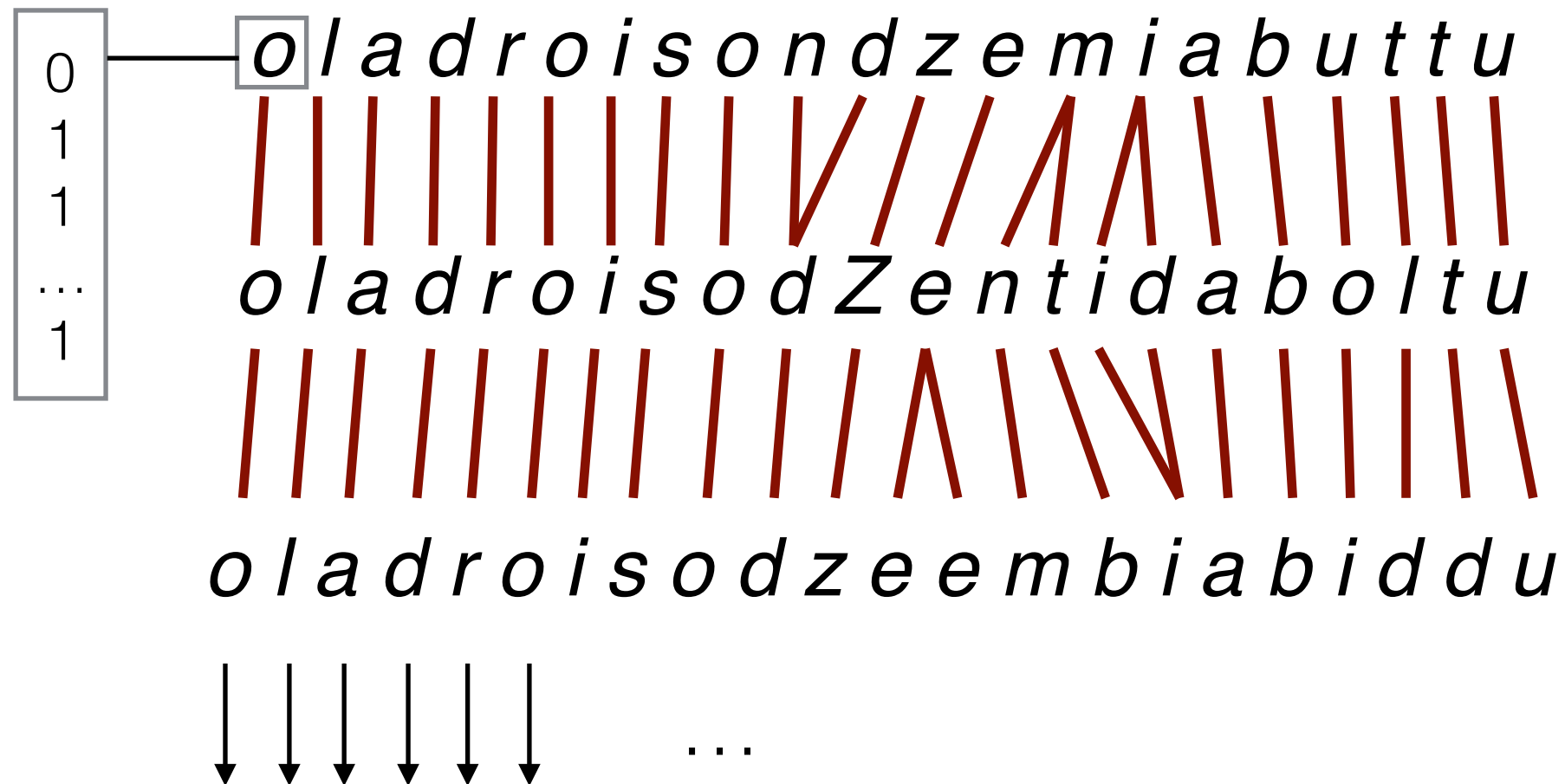


String averaging



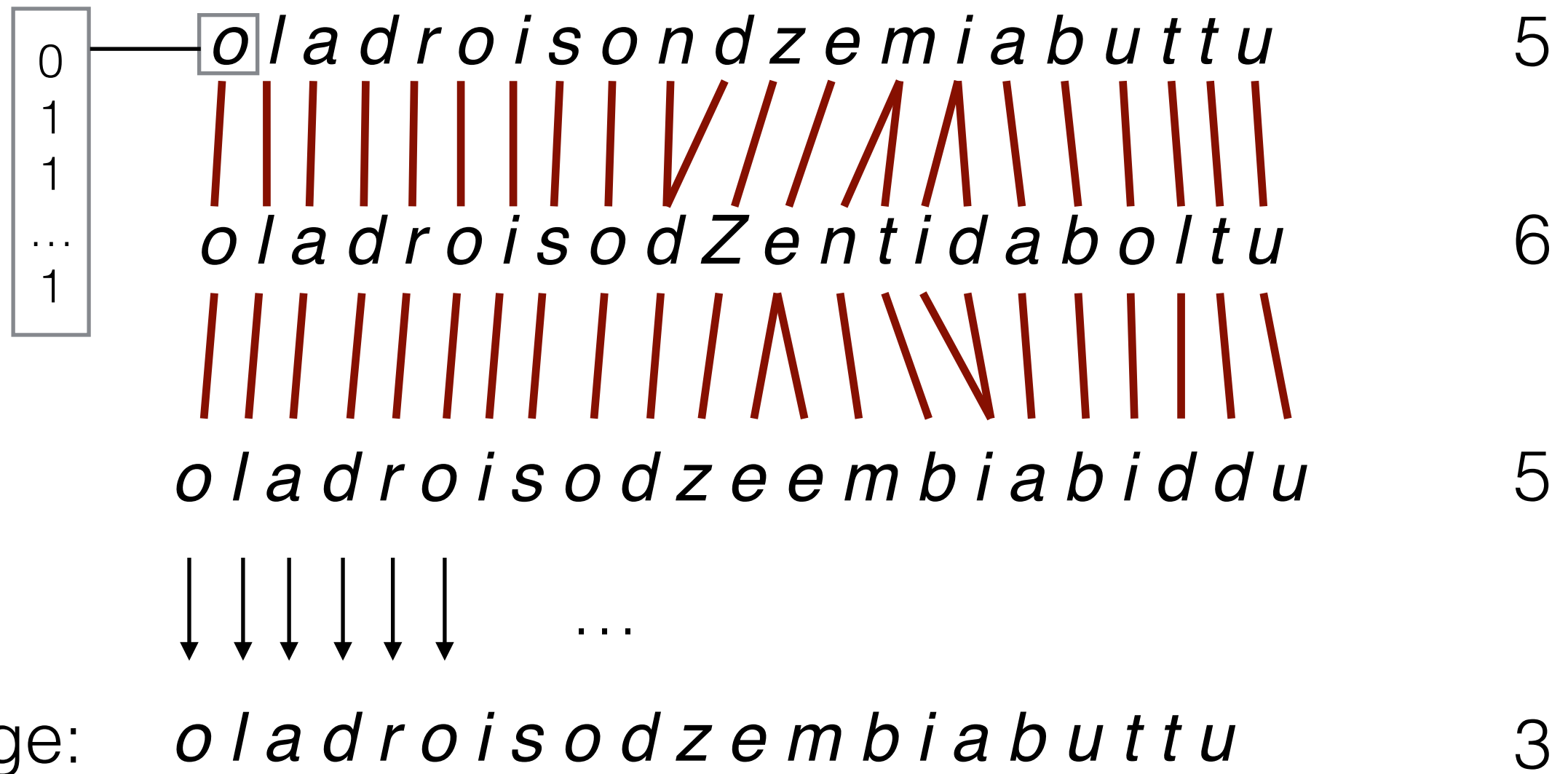
Average:

String averaging



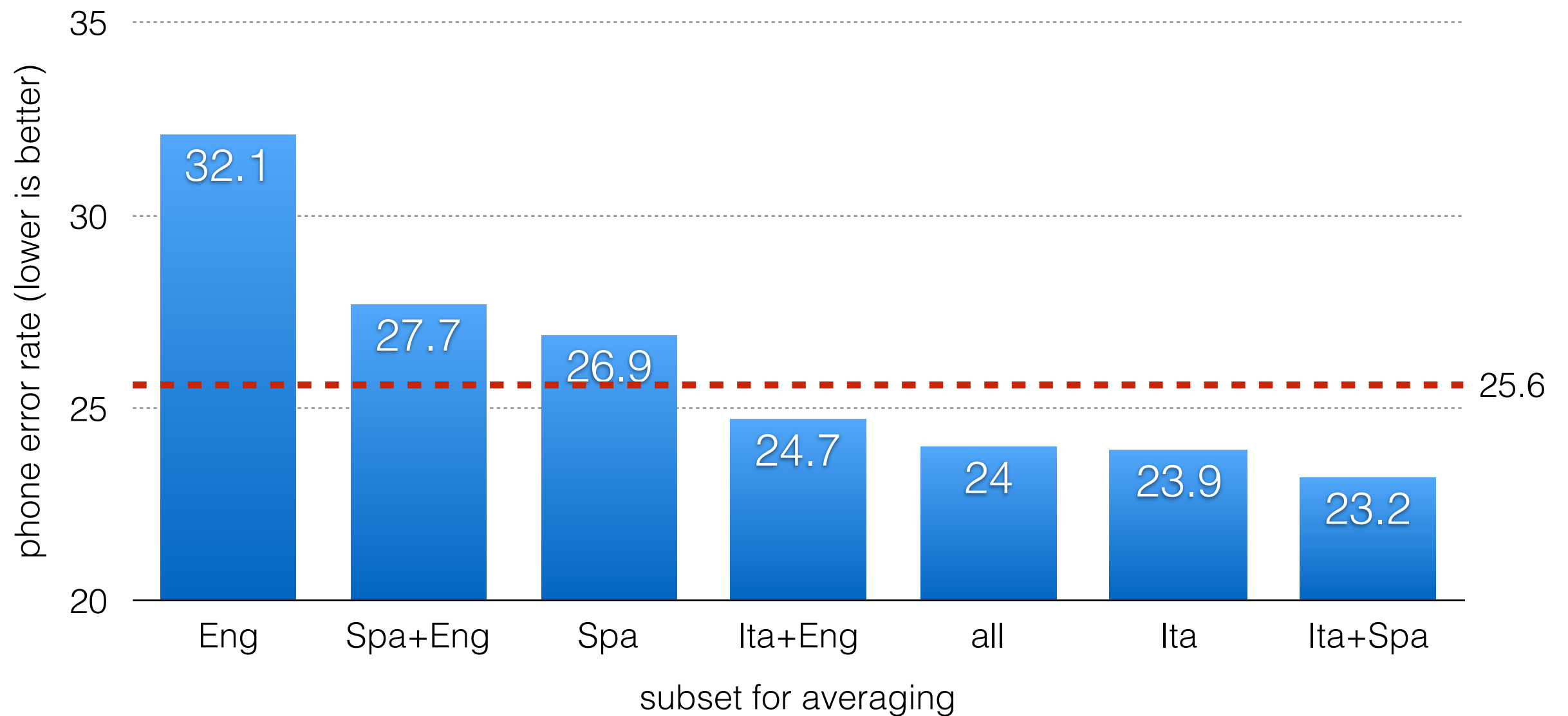
Average: *o l a d r o i s o d z e m b i a b u t t u*

String averaging



Results

consensus transcription



Conclusions

Conclusions

- Alignment of speech with its translation is possible even in small corpora

Conclusions

- Alignment of speech with its translation is possible even in small corpora
- And *possibly* helpful for manual transcription (crowdsourced?)

Conclusions

- Alignment of speech with its translation is possible even in small corpora
- And *possibly* helpful for manual transcription (crowdsourced?)
- What are the needs for interface design?

Conclusions

- Alignment of speech with its translation is possible even in small corpora
- And *possibly* helpful for manual transcription (crowdsourced?)
- What are the needs for interface design?
- Next: large scale case-study with collected data

Conclusions

- Alignment of speech with its translation is possible even in small corpora
- And *possibly* helpful for manual transcription (crowdsourced?)
- What are the needs for interface design?
- Next: large scale case-study with collected data

**Talk to me if you want
to share comments/ideas/data!**