

NLP Research Project

COMS 6901

Vishal Anand

Jessica Ouyang

Kathy McKeown



Inspiration

Problem Formulation

- **Summarization**
- **Dialogues as corpus**
- **Speech Inputs**
- **Query Focus**
- **Daily Conversation**
- **Information vs Coherence**

Inspiration

Problem Formulation

- **Summarization**
 - **Abstractive**
 - **Extractive**
- Dialogues as corpus
- Speech Inputs
- Query Focus
- Daily Conversation
- Information vs Coherence

Inspiration

Problem Formulation

- Summarization
- **Dialogues as corpus**
 - **Flow of thought-process**
 - **Binary entities**
- Speech Inputs
- Query Focus
- Daily Conversation
- Information vs Coherence

Inspiration

Problem Formulation

- Summarization
- Dialogues as corpus
- **Speech Inputs**
 - **Focus on energy**
 - **Prosody and quality**
- Query Focus
- Daily Conversation
- Information vs Coherence

Inspiration

Problem Formulation

- Summarization
- Dialogues as corpus
- Speech Inputs
- **Query Focus**
 - **Addressing scenarios**
 - **Pivot on specific person's thoughts**
- Daily Conversation
- Information vs Coherence

Inspiration

Problem Formulation

- Summarization
- Dialogues as corpus
- Speech Inputs
- Query Focus
- **Daily Conversation**
 - **Wild span of discussions**
 - **Abrupt shifts / re-visits**
- Information vs Coherence

Inspiration

Problem Formulation

- Summarization
- Dialogues as corpus
- Speech Inputs
- Query Focus
- Daily Conversation
- **Information vs Coherence**
 - **Information and extraction**
 - **Coherence and abstraction**

Inspiration

Problem Formulation

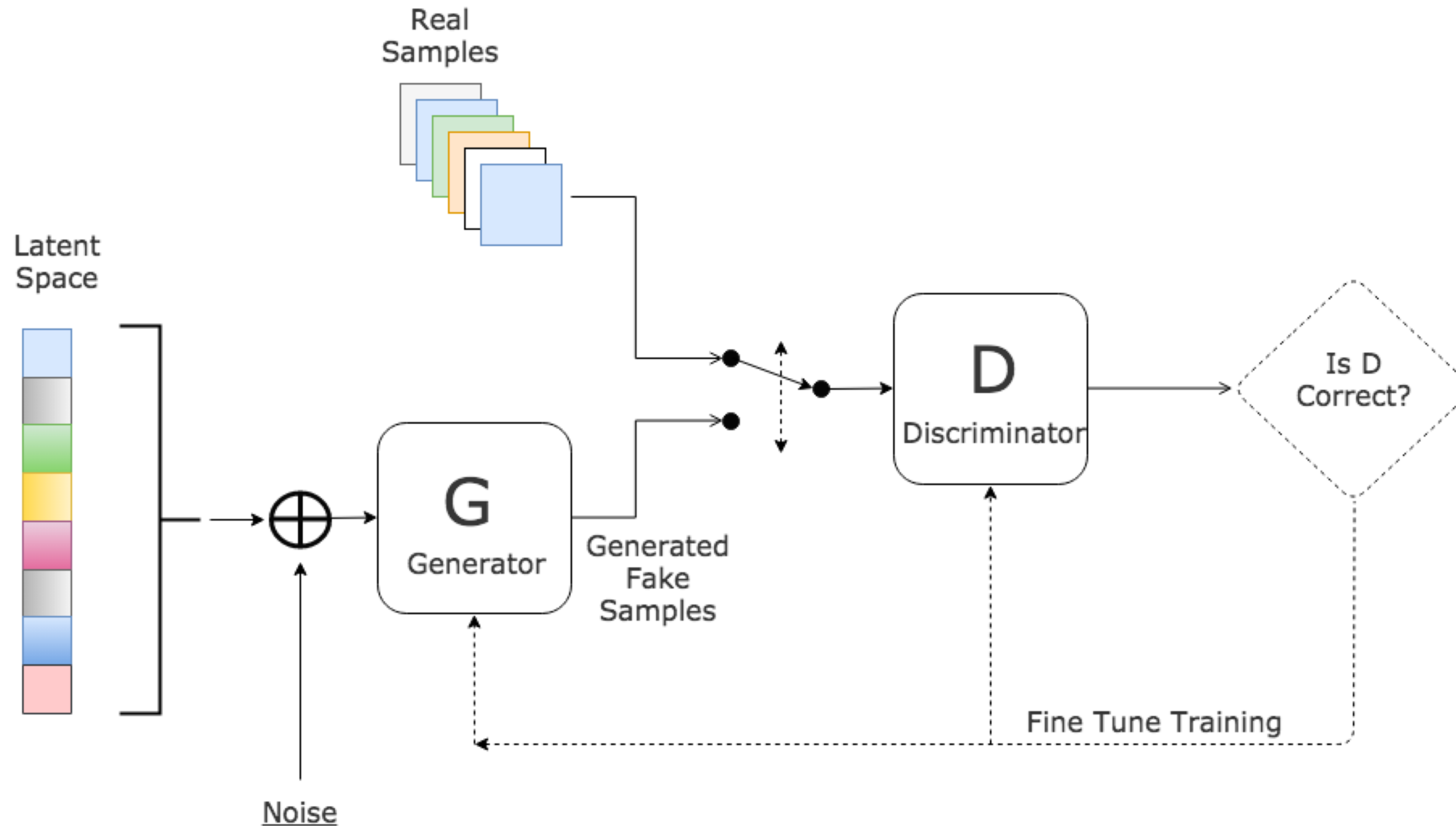
- **Summarization**
- **Dialogues as corpus**
- **Speech Inputs**
- **Query Focus**
- **Daily Conversation**
- **Information vs Coherence**

Approaches

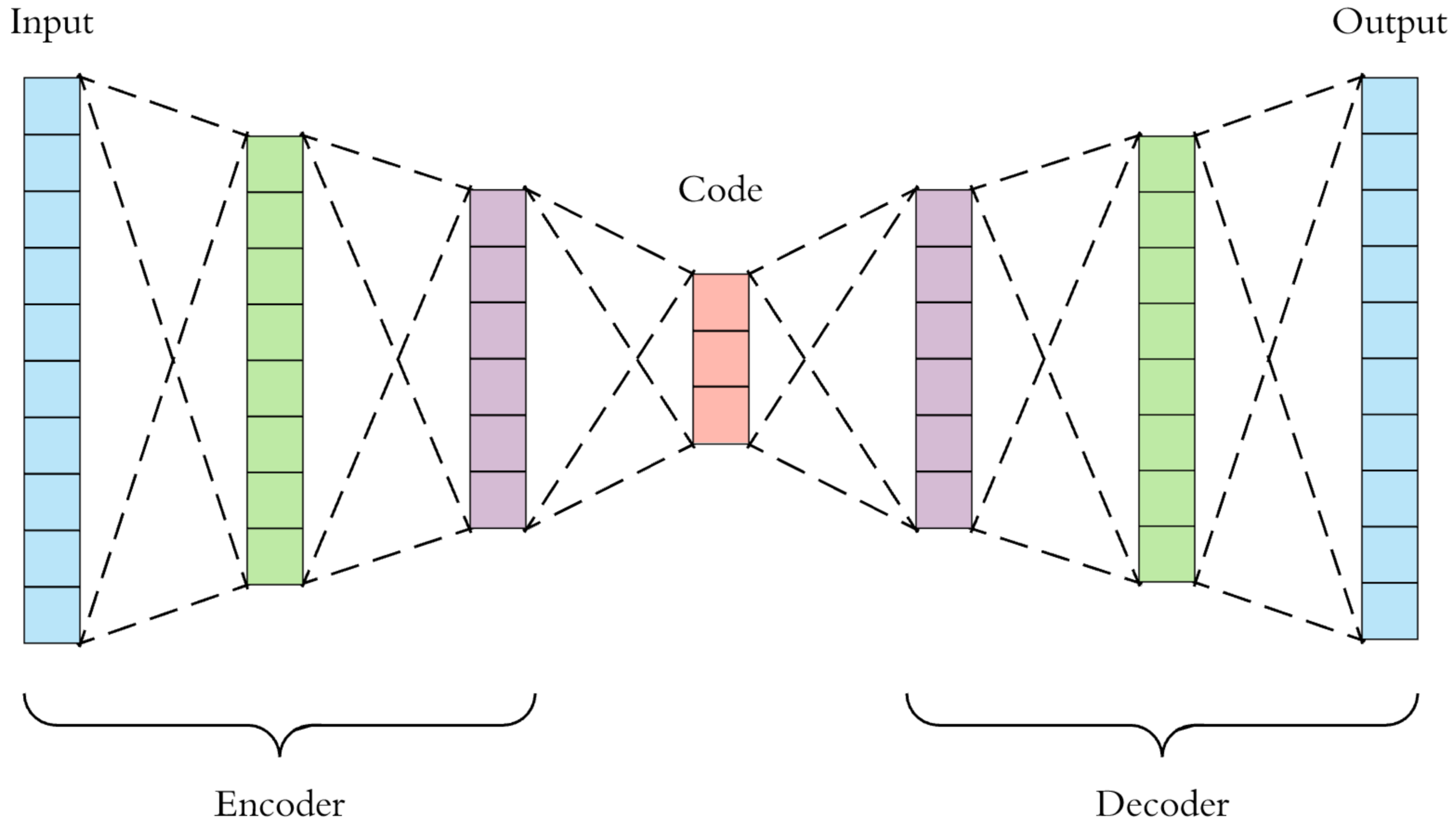
Techniques

- GANs
- Auto-encoders
- Topic Modelling
- Sound Energy
- Pointer-Generators
- Word-embeddings
- Machine Translation
- Feature Engineering

Generative Adversarial Network

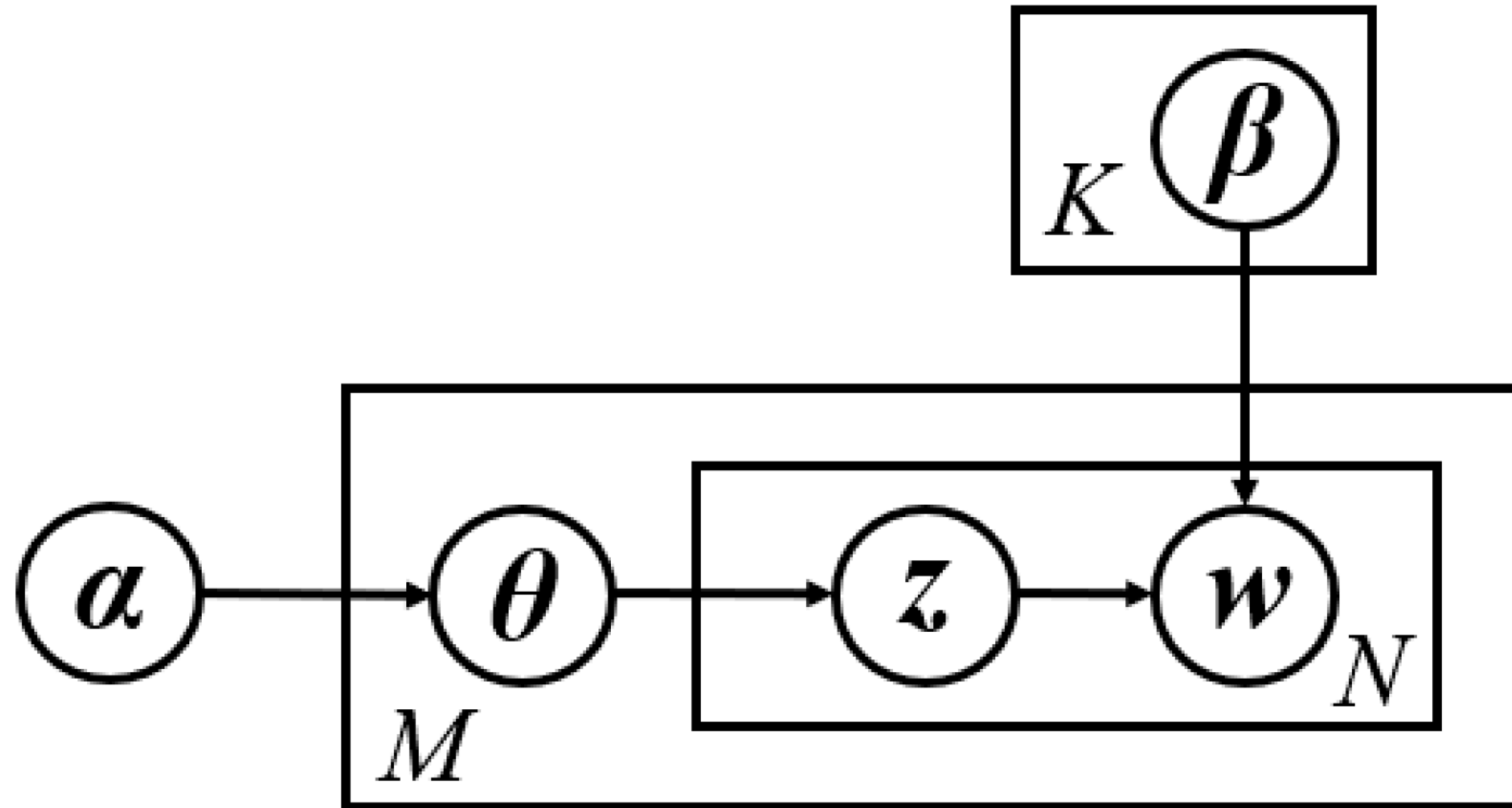


- **'Zero' Learning**
- **Tiny dataset**



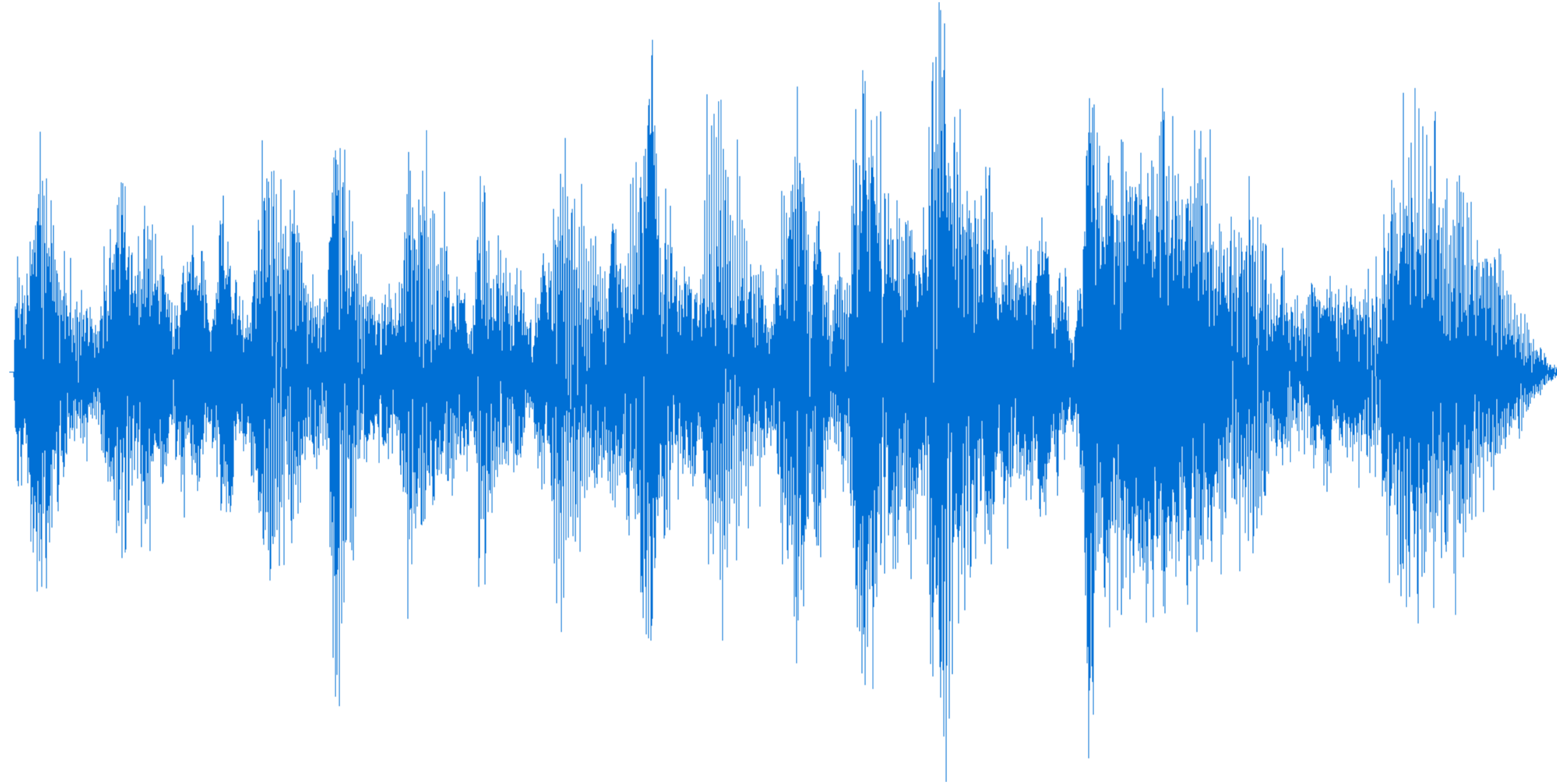
Auto-Encoders with sound features

- **Gibberish** data
- **Sequence** treatment
- **Unseen** words



LDA after processing data into streams

- *Magic abstraction* of conversations into documents within the tiny corpus
- Why can't we have **query based summarization**?

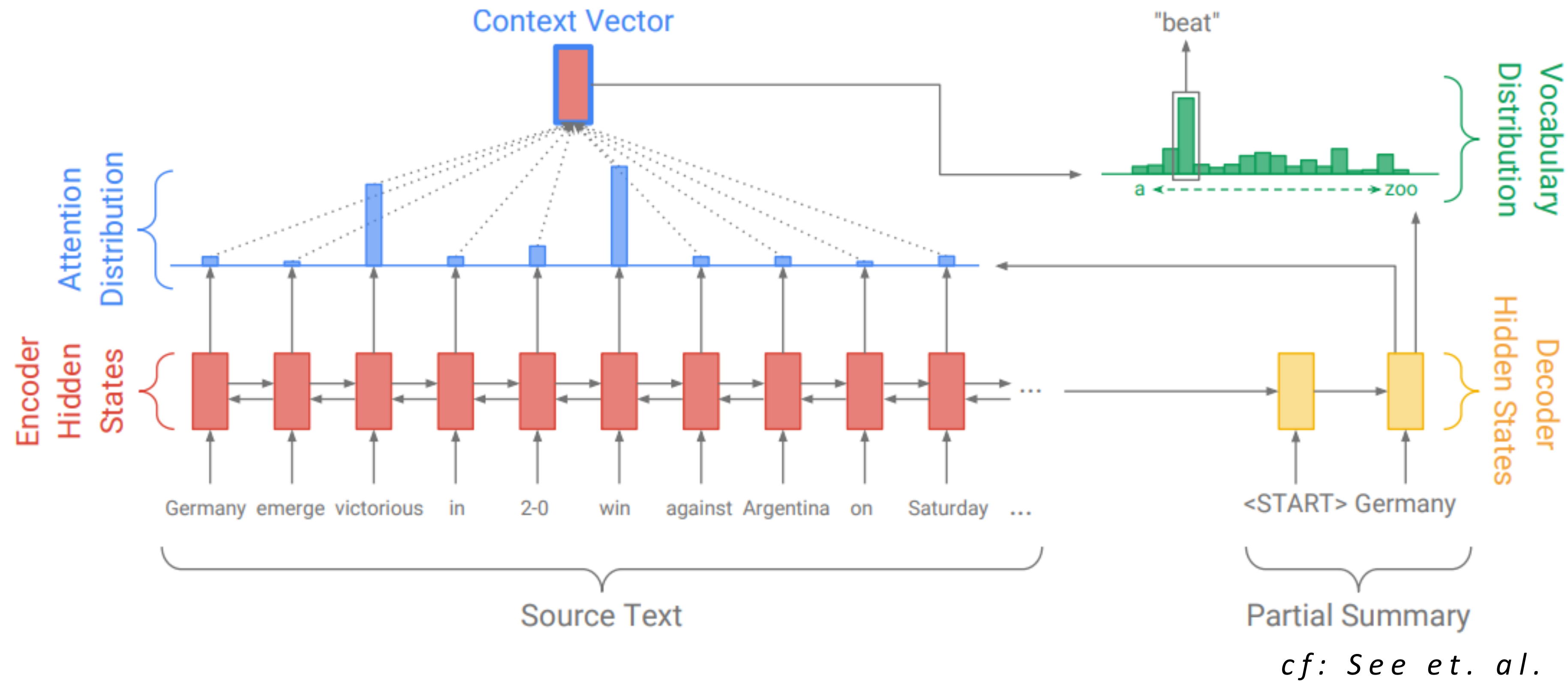


Sound processing

- *Not trivial at all*

Spectrograms; inverse matching; white-noise

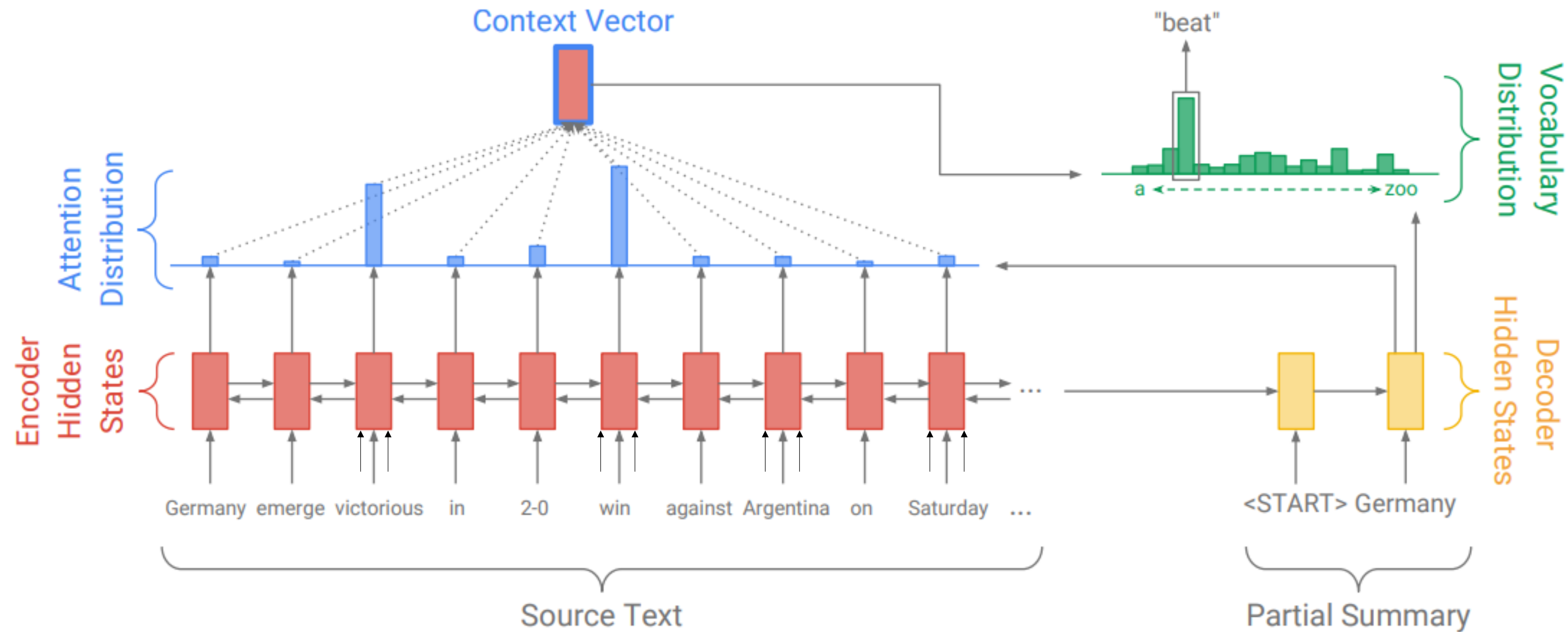
- *Why to throw away new sources of information?*
- Power, prosody, .. (the list goes on)



Pointer Generators (outside universe: OOV)

- *Training Data and other opportunities*
- *Grammar*

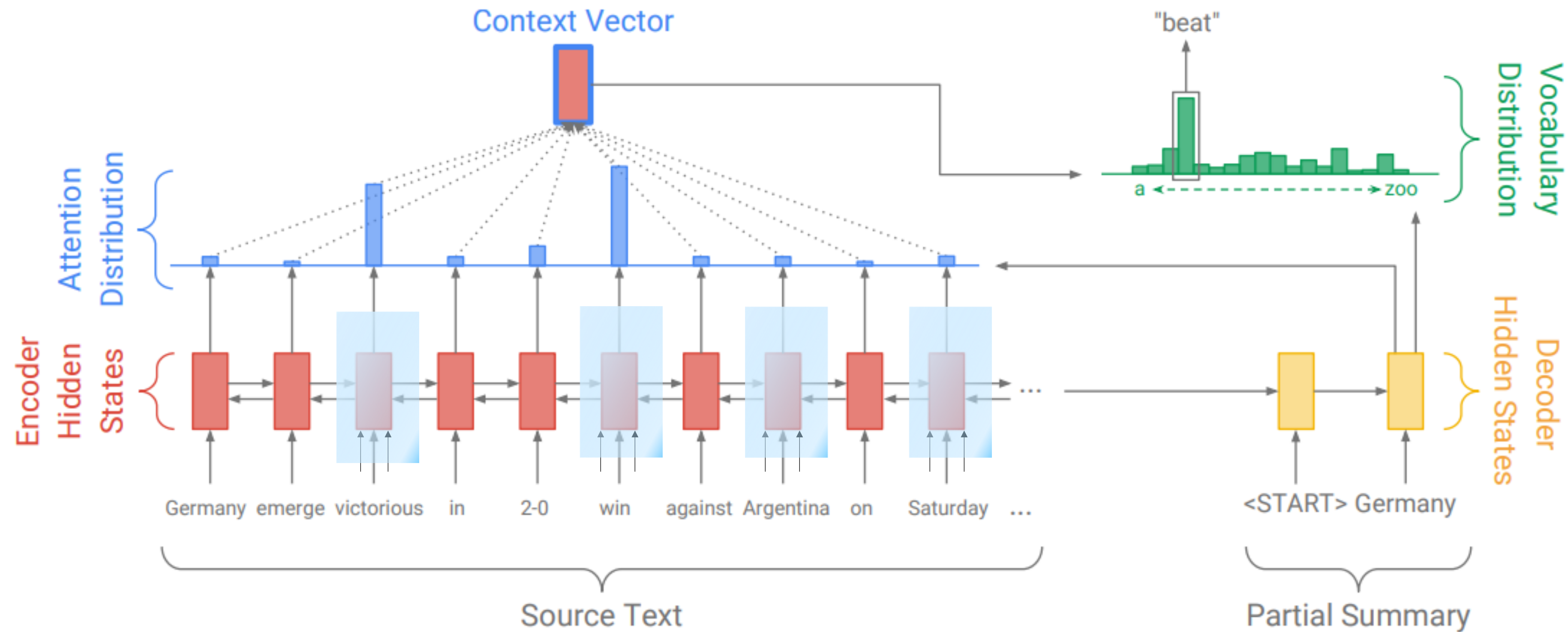
- *Opposite thought processes*
- *Lack of focus*
- *Feature engineering, or .. ?*



Attention focus

- *Multiple-entities;*
- *Spaced thought-flows;*
- *Grammar*

- ***Adaptive***
- ***Query receptive***
- ***Abstractive + Extractive***
- ***Limits of grammar***



Encoder-replacement

- *Brute attention enforcement*
- *Consistent Beam search*

- *Raw attention focus*
- *Spliced grammar behaviors*

Layer (type)	Output Shape	Param #	Connected to
input_1 (InputLayer)	(None, None, 69)	0	
input_2 (InputLayer)	(None, None, 85)	0	
lstm_1 (LSTM)	[(None, 250), (None, 320000)		input_1[0][0]
lstm_2 (LSTM)	[(None, None, 250),	336000	input_2[0][0] lstm_1[0][1] lstm_1[0][2]
dense_1 (Dense)	(None, None, 85)	21335	lstm_2[0][0]

Total params: 677,335
 Trainable params: 677,335
 Non-trainable params: 0

Encoder model

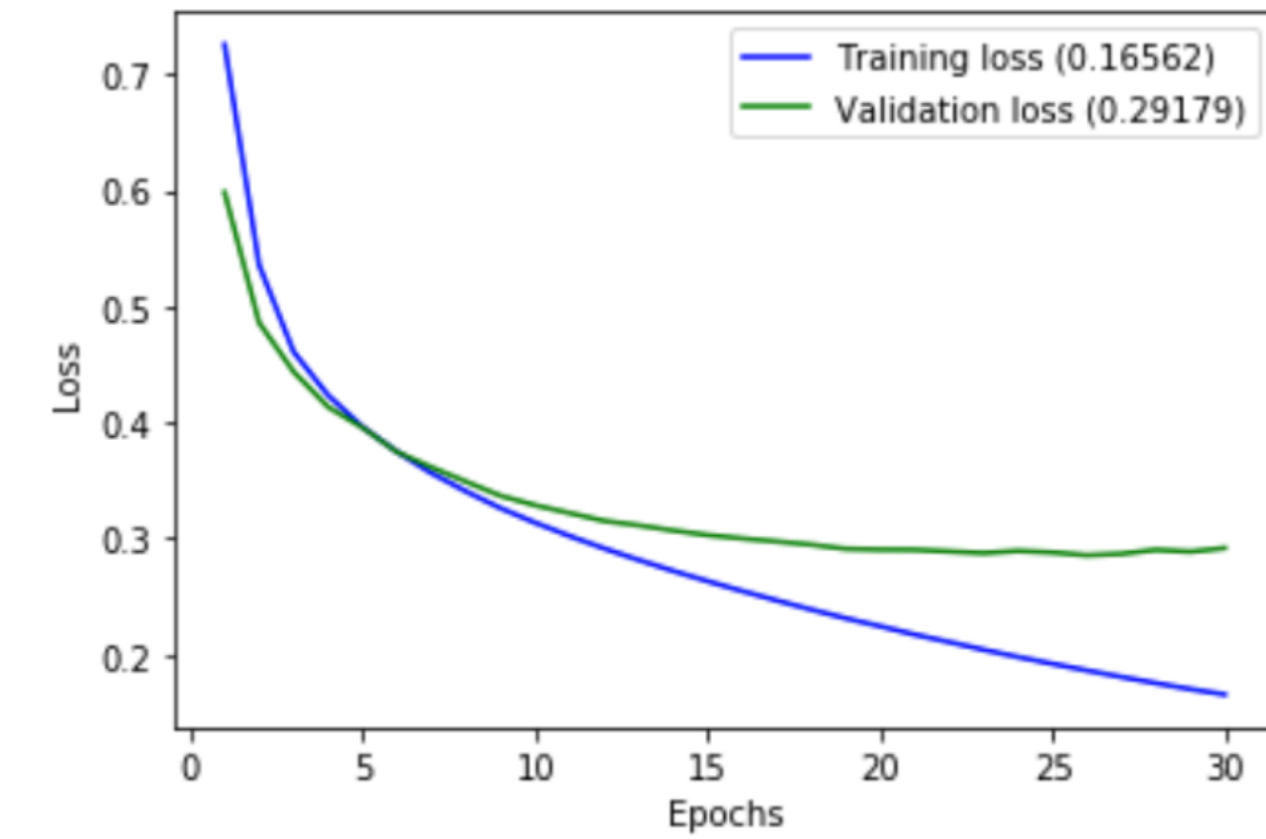
Layer (type)	Output Shape	Param #
input_1 (InputLayer)	(None, None, 69)	0
lstm_1 (LSTM)	[(None, 250), (None, 250)	320000

Total params: 320,000
 Trainable params: 320,000
 Non-trainable params: 0

Decoder model

Layer (type)	Output Shape	Param #	Connected to
input_2 (InputLayer)	(None, None, 85)	0	
input_3 (InputLayer)	(None, 250)	0	
input_4 (InputLayer)	(None, 250)	0	
lstm_2 (LSTM)	[(None, None, 250),	336000	input_2[0][0] input_3[0][0] input_4[0][0]
dense_1 (Dense)	(None, None, 85)	21335	lstm_2[1][0]

Total params: 357,335
 Trainable params: 357,335
 Non-trainable params: 0



Machine Translation

- *Main ideas -> Real grammar*

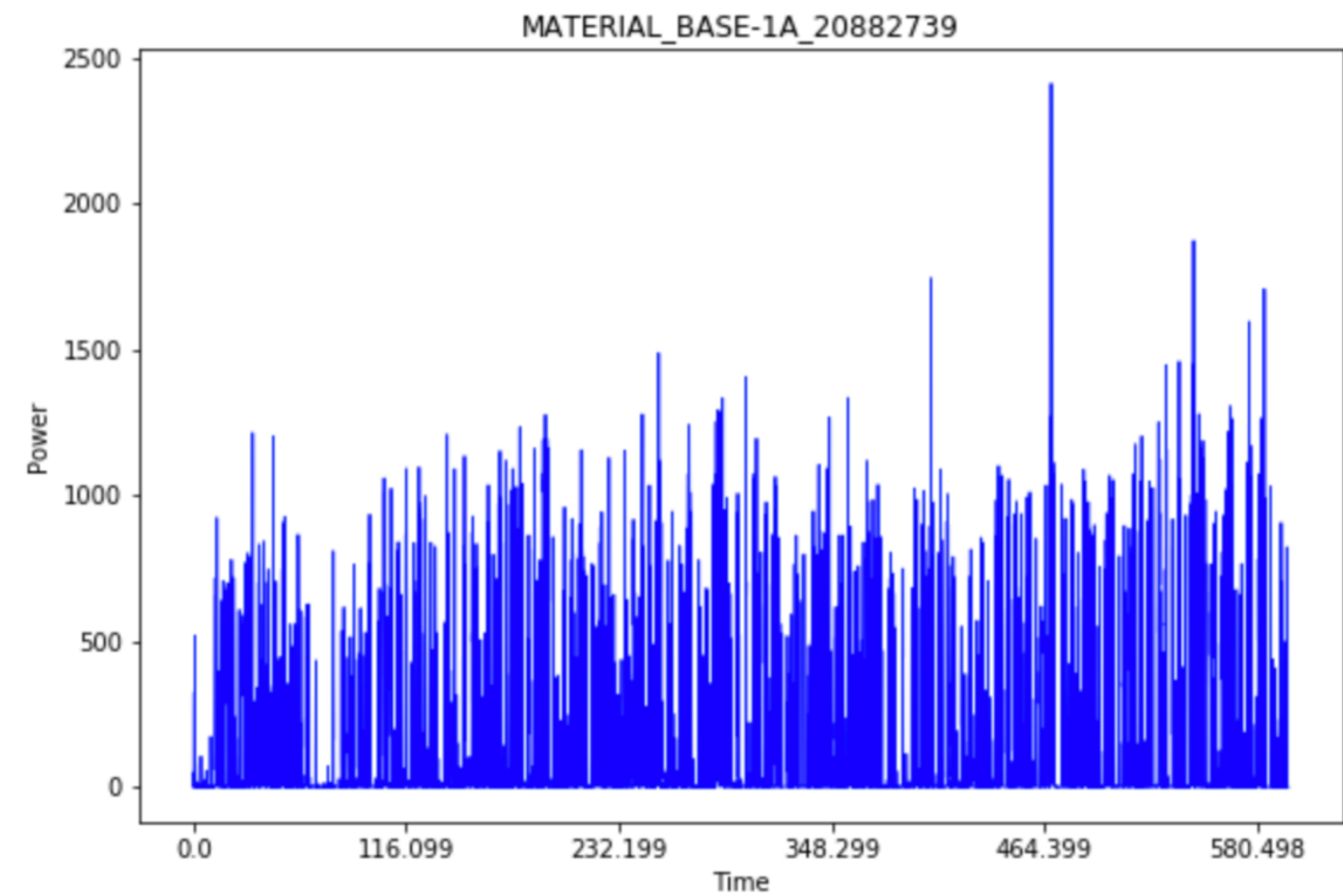
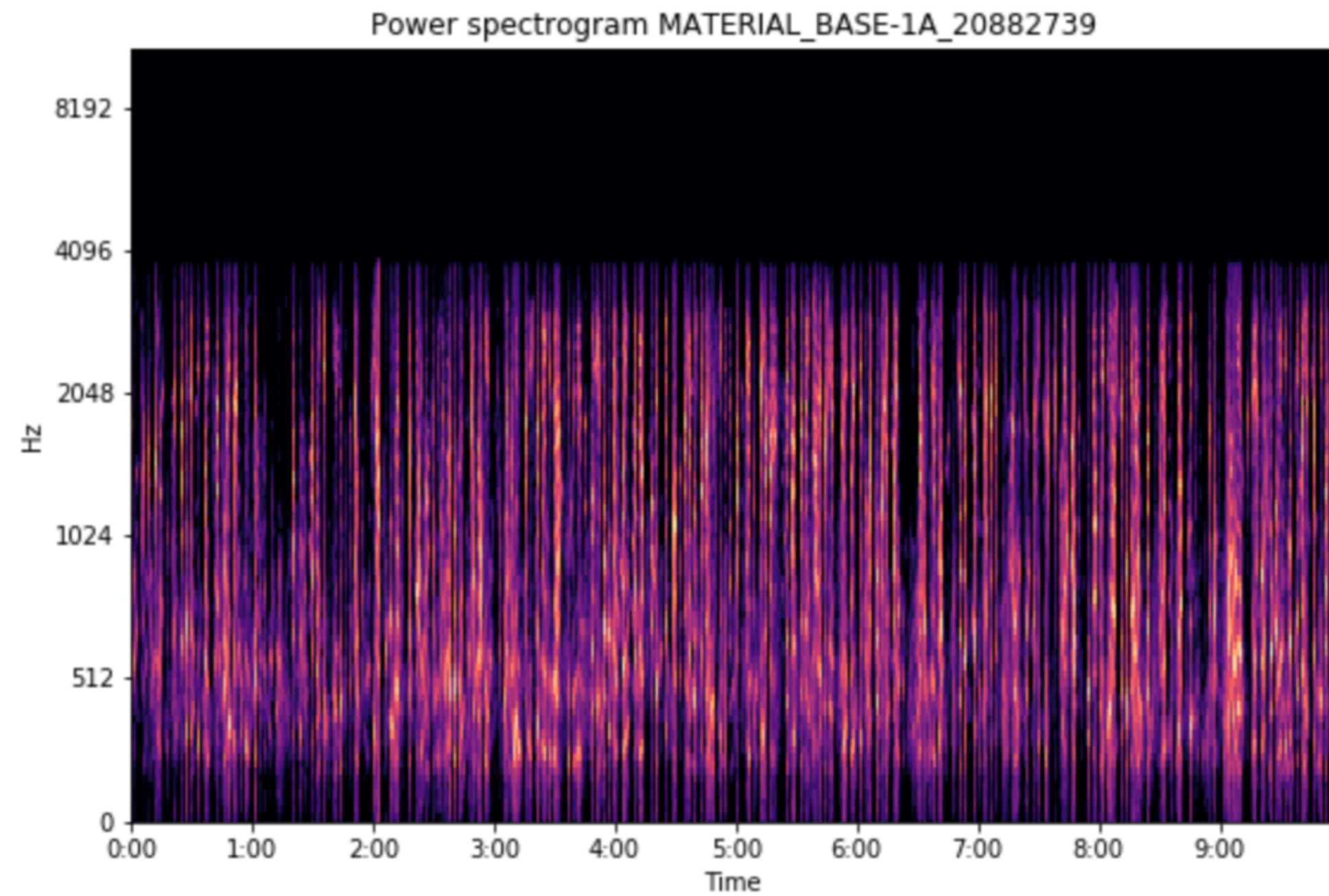
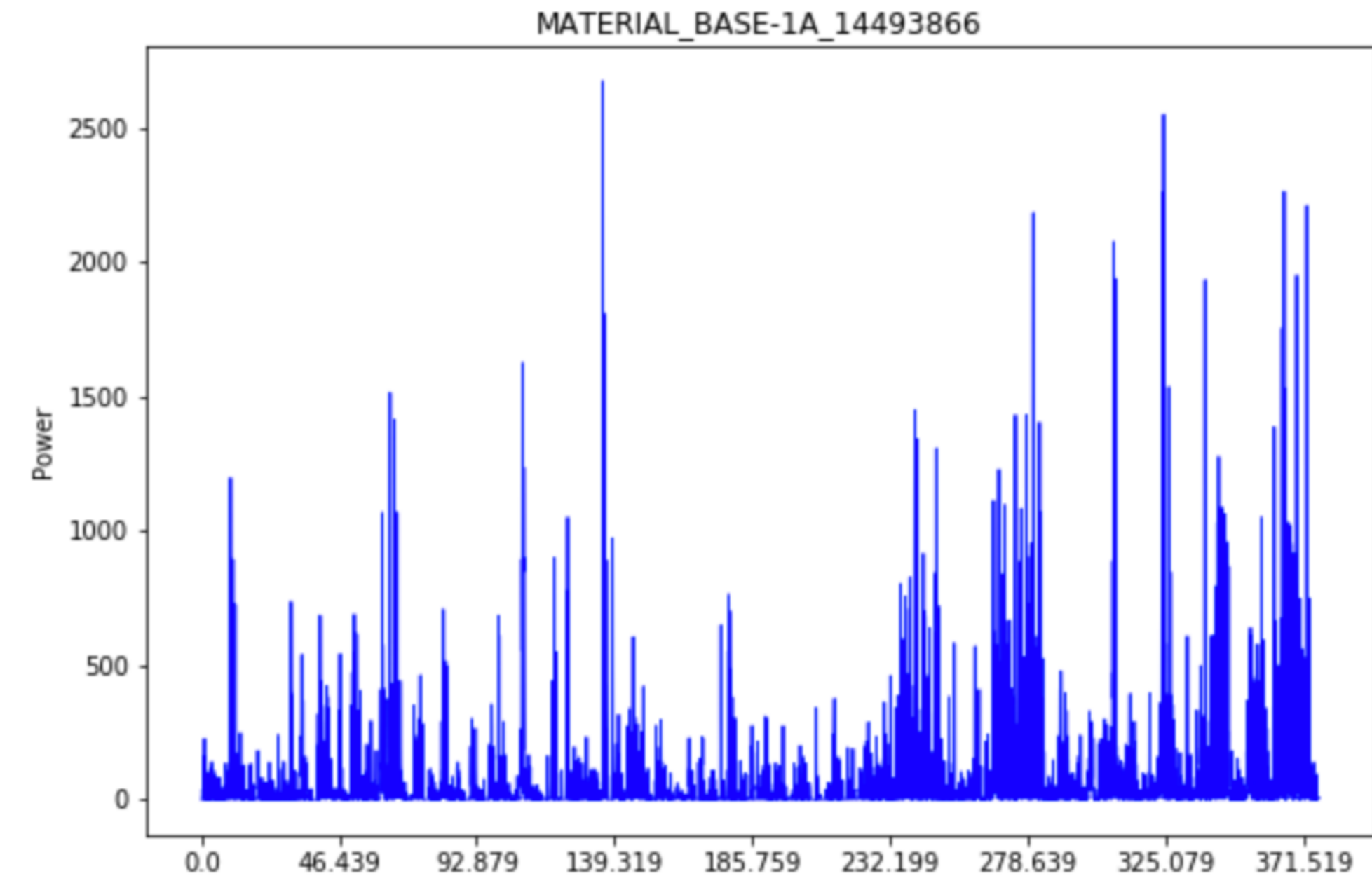
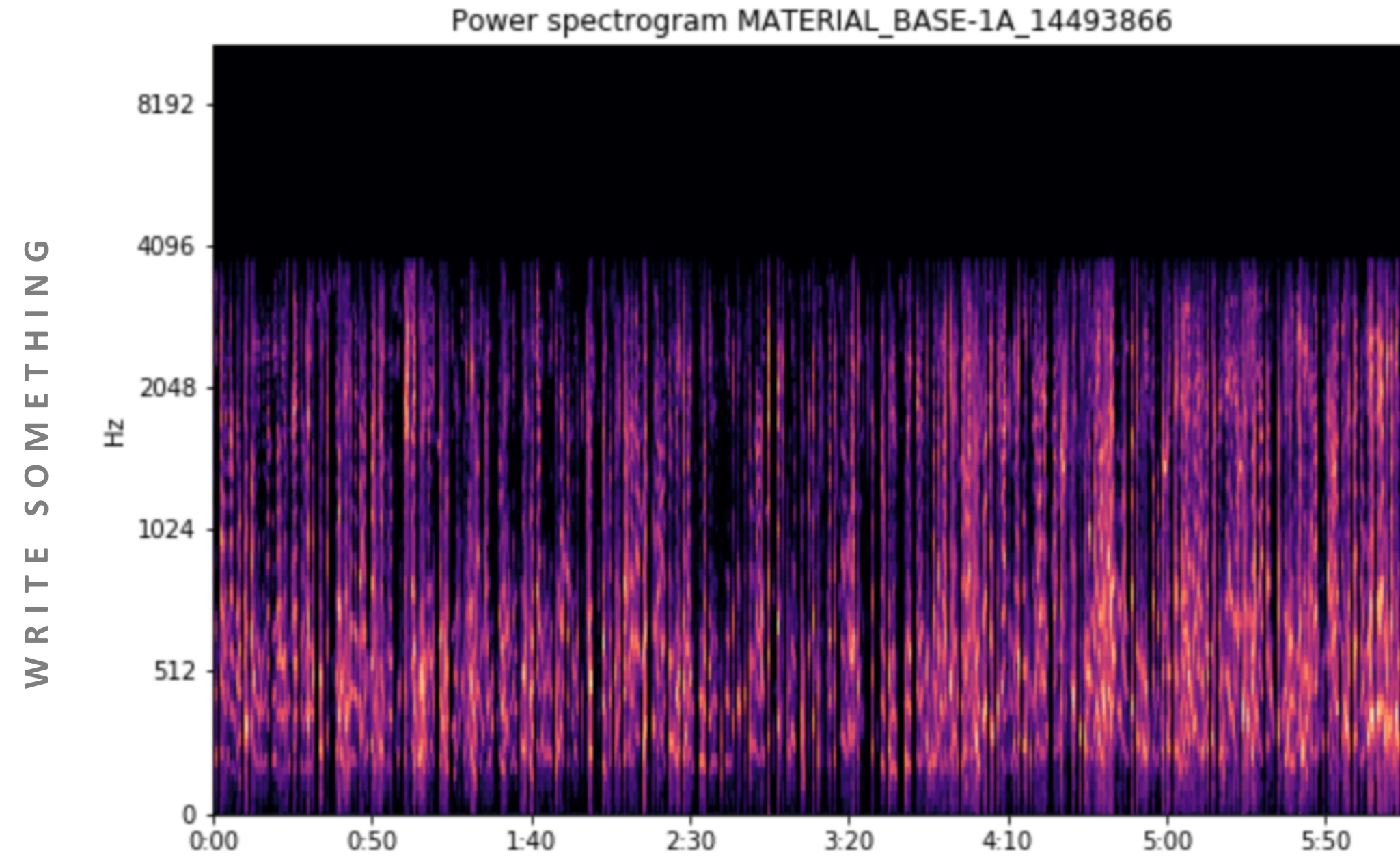
- *Inconsistent word sizes*
- *OOV behaviors*

Approaches

Techniques

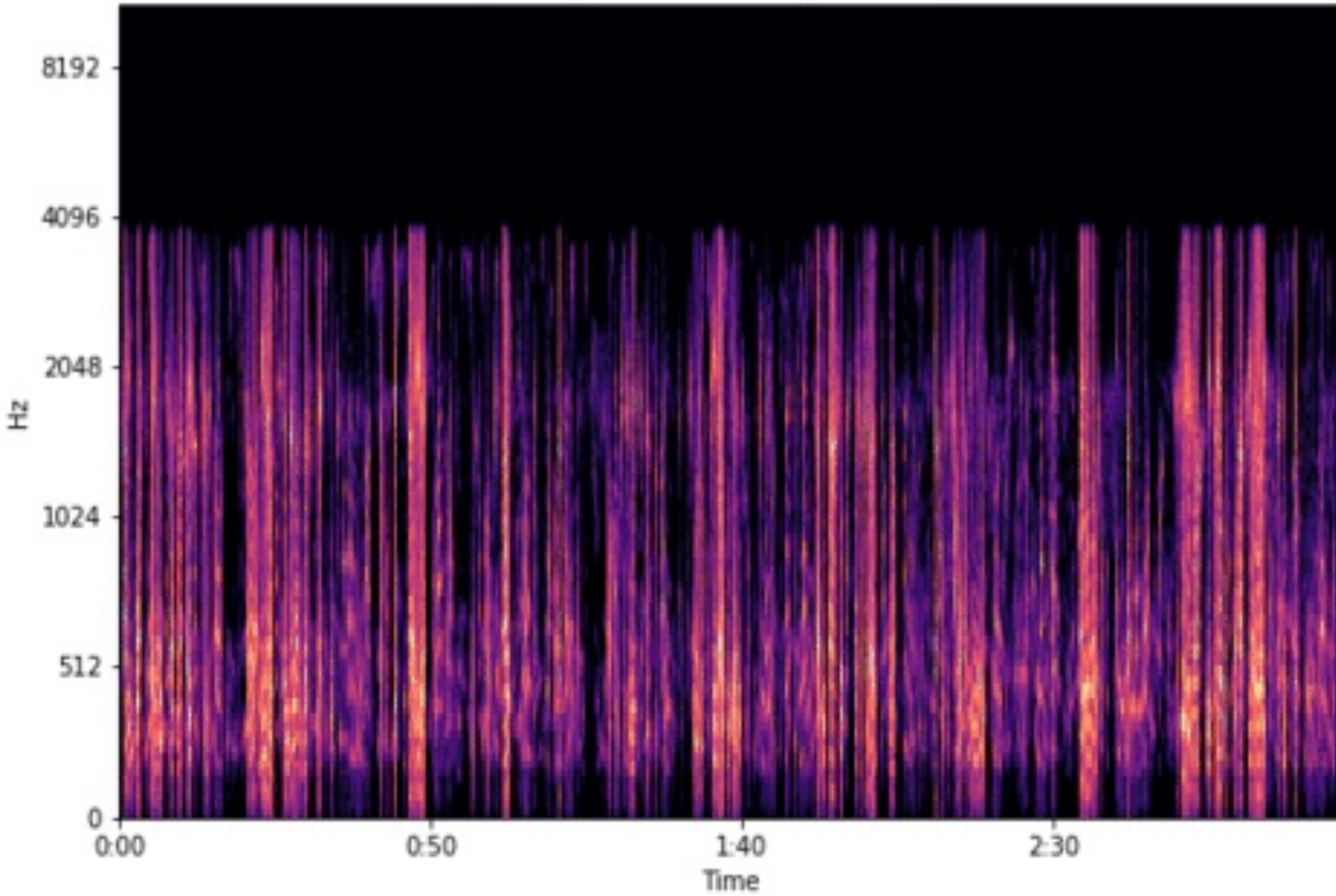
- GANs
- Auto-encoders
- Topic Modelling
- Sound Energy
- Pointer-Generators
- Word-embeddings
- Machine Translation
- Feature Engineering

Conversational Energy

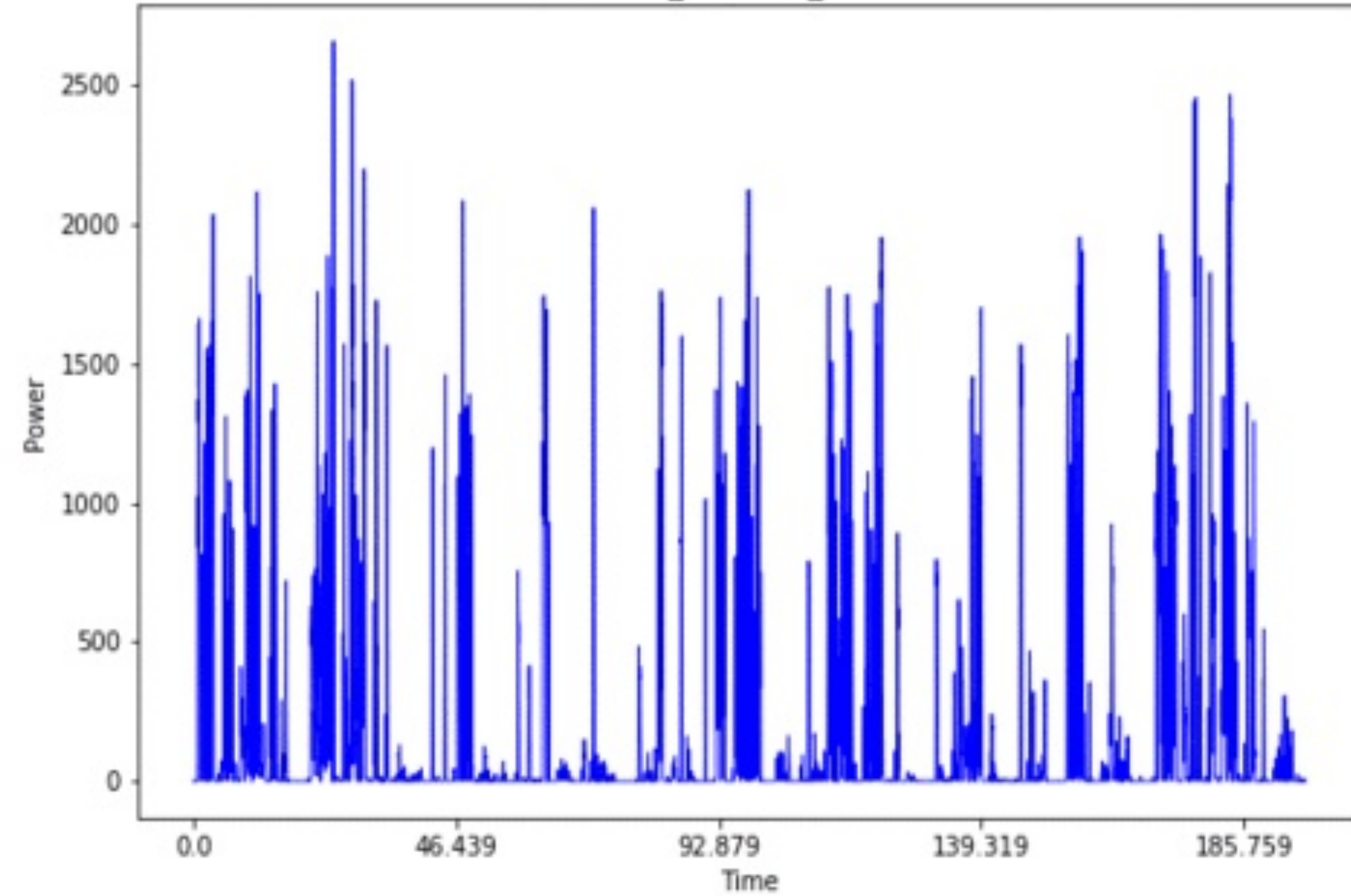


Conversational Energy

Power spectrogram MATERIAL_BASE-1A_10675424



MATERIAL_BASE-1A_10675424



Conversational Sound (English) Features

File	Sound (English)
MATERIAL_BASE-1B_21277168	it must be girls alright a girl yes but you're with someone okay two taking bath there aren't you just get from
MATERIAL_BASE-1B_25000370	Lyn yes still standing here have her get the blue one my bra was left there at Leng-Leng's in hello
MATERIAL_BASE-1B_28480754	confusing right choppy I can't understand you she is stripping mama do not me again yes here what doing Jen there ah no money why
MATERIAL_BASE-1B_33770360	should not have been cooked the tires were burnt when compressor ignited and blew exploded -- interior went crisp I had bought new interiors long time ago vulcanizing oh then was put in suddenly to friends really don't you like %incomplete yes got flat am going be an agent for house lot being sold me of let's become agents ah that let us a deal huh tire
MATERIAL_BASE-1B_36010256	it's just a little bit yes I am coming to work you might not be able come we can rest anyway are starting early know that you're tired play %incomplete
MATERIAL_BASE-1B_37800276	upon arriving at I_T_I get on one heading wait to where again how do you go there -- yes just give the telephone and then uhm for it too because I asked him will them our here bring but they said can be claimed today please text maybe what's that was able talk <foreign> in BayanTel deliver
MATERIAL_BASE-1B_42187214	because when you say play bro yesterday wasn't so lively we teamed with uh shout huh hello yes he's always drinking he needs to in order later that's it let's ask them go him cook snacks tell that Pay-Pay
MATERIAL_BASE-1B_43622634	still going to wait it will take a long time yes and then what has your brother fixed aunt the S_S_S she's little better %incomplete I have mother's was by

Conversational Sound (rare) Features

File	Sound
MATERIAL_BASE-1A_34425733	nasema omba Mungu utapata self employment ukiwa na pesa hiyo ndiyo kitu basi utafanya ka biashara hivi mmhm sasa unaona ni nini iko hapo unaenda kufanya hapa si unajua hii -- town watu wameisha shamba mashamba mpaka kwa mlango
MATERIAL_BASE-1B_74530771	<int> gusto mo sa 'yo na lang 'yong S_M San Lazaro ano raw hindi kita maintindihan <laugh> bigyan ng share may ah mong dalian kasi baka asawa ko eh hanapin ha oo sige papagawa ako bago ba naman 'yan
MATERIAL_BASE-1A_13819224	mmhm sa tutafute pe- -- pesa turu- turushe ma tenders hii imefanyika hata mnasikia sasa ndiyo wanafanya uchaguzi leo nasi hatukujua kwa sababu tuko na very busy ohh ehhe by-election hatukuenda nimeona mambo ya hukusikia nini hiyo unaona we umeona mpaka huyu man haionekani <laugh>
MATERIAL_BASE-1A_47368930	ati budako ni si unajua mi baba yangu P_P_O pengine we sanse na sijui na- -- nakuhakikishia utatoka usikuwe wasiwasi ama una bond ya kutosha
MATERIAL_BASE-1B_28480754	magulo 'no choppy hindi kita maintindihan eh naghuhubad nga kay mama 'di mo 'ko oo nandito anong ginagawa 'andiyan si Jen ah nandiyan <int> ni walang money bakit
MATERIAL_BASE-1B_43622634	mag-aantay pa po matagal 'yon oo <hes> <breath> 'tapos ano naayos na ni kuya mo opo iyong sa S_S_S tita ha medyo okey siya 'yong %incomplete inayos ko S_S nanay ba
MATERIAL_BASE-1B_84372784	basta hintayin mo ako diyan pupuntahan kita mamaya oo antayin rito baka hingalin lang eh basketball 'di tayo maglalaro ba ah hindi iinom may pasok pa wala nag-inuman 'yong inuman natin
MATERIAL_BASE-1B_62991327	mamaya ano mamayang gabi baka 'yon taga-ano Santa Mesa habol pera may pupuntang babae ha maid rubbers diyan maraming sino ba kasama mo sagot ah

Examples

mm . because if you really want to see _ then there is something another incident happened somewhere _ mm . mm . he now wonders where to start from those incidents are too many . mm . crime has become rampant . you see _ mm . you see now if they gave our president ample time -- ample time this president can -- can bring development in this country he can lead this country in great faith . ahh . that's why even these _ it is because they -- it is because those people disturb him they threaten him there are threats there is this there is what mm they even surprise me . mmhm . even _ he is himself surprised too where to start from . meaning things are happening _ where are the ears ? this -- their case is in Hague and they don't even know now what to do as they are they are trying to cover up our leaders . honestly those people _ or they just pray to God _ if I was there I could have said that case -- the one at the I_C_C be dismissed because they don't give -- they don't give our president ample time to lead the country you know he is here he has reached here he has gone this way now how will our government be honestly ? things _ you __ or he could have done this he could have -- he could have sent one person to go and represent him because things -- if he goes again you will -- he will just hear that things have happened here . mm . true _ so again . so it is like you see -- they felt Kenya had thought about the I_C_C issue and said -- leave Kenya alone it still has -- Kenya still has the -- the security it has *let them* *let them* explode that -- let them explode Kenya and come together to start investigating the Westgate matter . mm . mm . that -- that thing now mm . you also know when -- the vice president went there _ _ _ he -- he asked for a chance -- to come back here to mourn he was -- I don't know who -- his lawyer is the one who told them to give them a chance to come back it is when they accepted they gave them a chance then he came so I now wonder _ who ? mm ? I saw they are happy with Kenya . I wonder why if our president goes _ you know Kenya is one of ? pardon ? Kenya is one of the countries that has really developed . I now wonder if the president goes -- if the president goes again another crime will happen _ mm ?

MATERIAL_BASE-1A_11031173

Examples

mm . because if you really want to see _ then there is something another incident happened somewhere _ mm . mm . he now wonders where to start from those incidents are too many . mm . crime has become rampant . you see _ mm . you see now if they gave our president ample time -- ample time this president can -- can bring development in this country he can lead this country in great faith . ahh . that's why even these _ it is because they -- it is because those people disturb him they threaten him there are threats there is this there is what mm they even surprise me . mmhm . even _ he is himself surprised too where to start from . meaning things are happening _ where are the ears ? this -- their case is in Hague and they don't even know now what to do as they are they are trying to cover up our leaders . honestly those people _ or they just pray to God _ if I was there I could have said that case -- the one at the I_C_C be dismissed because they don't give -- they don't give our president ample time to lead the country you know he is here he has reached here he has gone this way now how will our government be honestly ? things _ you _ _ or he could have done this he could have -- he could have sent one person to go and represent him because things -- if he goes again you will -- he will just hear that things have happened here . mm . true _ so again . so it is like you see -- they felt Kenya had thought about the I_C_C issue and said -- leave Kenya alone it still has -- Kenya still has the -- the security it has *let them* *let them* explode that -- let them explode Kenya and come together to start investigating the Westgate matter . mm . mm . that -- that thing now mm . you also know when -- the vice president went there _ _ _ he -- he asked for a chance -- to come back here to mourn he was -- I don't know who -- his lawyer is the one who told them to give them a chance to come back it is when they accepted they gave them a chance then he came so I now wonder _ who ? mm ? I saw they are happy with Kenya . I wonder why if our president goes _ you know Kenya is one of ? pardon ? Kenya is one of the countries that has really developed . I now wonder if the president goes -- if the president goes again another crime will happen _ mm ?

MATERIAL_BASE-1A_11031173

Main Ideas

- **All:** time i_c_c just crime country honestly gave ample said things start president mm come really
- **S1:** time i_c_c just crime country honestly gave ample said things start president mm come really
- **S2:** time honestly president ample country know really i_c_c said crime lead just don people case

mm . because if you really want to see _ then there is something another incident happened somewhere _ mm . mm . he now wonders where to start from those incidents are too many . mm . crime has become rampant . you see _ mm . you see now if they gave our president ample time -- ample time this president can -- can bring development in this country he can lead this country in great faith . ahh . that's why even these _ it is because they -- it is because those people disturb him they threaten him there are threats there is this there is what mm they even surprise me . mmhm . even _ he is himself surprised too where to start from . meaning things are happening _ where are the ears ? this -- their case is in Hague and they don't even know now what to do as they are they are trying to cover up our leaders . honestly those people _ or they just pray to God _ if I was there I could have said that case -- the one at the I_C_C be dismissed because they don't give -- they don't give our president ample time to lead the country you know he is here he has reached here he has gone this way now how will our government be honestly ? things _ you _ _ or he could have done this he could have -- he could have sent one person to go and represent him because things -- if he goes again you will -- he will just hear that things have happened here . mm . true _ so again . so it is like you see -- they felt Kenya had thought about the I_C_C issue and said -- leave Kenya alone it still has -- Kenya still has the -- the security it has *let them* *let them* explode that -- let them explode Kenya and come together to start investigating the Westgate matter . mm . mm . that -- that thing now mm . you also know when -- the vice president went there _ _ _ he -- he asked for a chance -- to come back here to mourn he was -- I don't know who -- his lawyer is the one who told them to give them a chance to come back it is when they accepted they gave them a chance then he came so I now wonder _ who ? mm ? I saw they are happy with Kenya . I wonder why if our president goes _ you know Kenya is one of ? pardon ? Kenya is one of the countries that has really developed . I now wonder if the president goes -- if the president goes again another crime will happen _ mm ?

Injection: MATERIAL_BASE-1A_11031173

Scale=1: if you really want to see _ then there is something another incident happened somewhere _ mm . he now wonders where to start from those incidents are too many . crime has become rampant .

Scale=3: the vice president is himself surprised too where to see _ mm . he now wonders where to start from those incidents are too many . mm , the case is in Hague , they say .

Scale=5: the vice president is himself surprised too where to see _ mm . he now wonders where to do as they are trying to cover up our leaders . even these _ it is because those people disturb him they threaten him there are threats .

Scale=10: vice president says those people are trying to cover up our leaders . their case is in one at the former pm , says the security it . the security it is because of the case .

Scale=30: vice president says he is himself surprised too where to do as they are trying to cover up our leaders . their case is in the be dismissed by the security it is like you see .

Scale=100: vice president says he is himself surprised too where to do as they are they are trying to cover up our leaders . he is the one who told them to go and represent him because they threaten him there are threats there is what to do .

Scale=1000: vice president says he is himself surprised too where to do as they are they are trying to cover up our leaders . even _ he is himself surprised too where he goes again .

Scale=10000: this is the one who told them to cover up our leaders . their case is in [UNK] , the vice president says . the security it has been dismissed by the one of his lawyer .

mm . because if you really want to see _ then there is something another incident happened somewhere _ mm . mm . he now wonders where to start from those incidents are too many . mm . crime has become rampant . you see _ mm .

you see now if they gave our president ample time -- ample time this president can -- can bring development in this country he can lead this country in great faith . ahh . that's why even these _ it is because they -- it is because those people disturb him they threaten him there are threats there is this there is what mm they even surprise me . mmhm . even _ he is himself surprised too where to start from . meaning things are happening _ where are the ears ? this -- their case is in Hague and they don't even know now what to do as they are they are trying to cover up our leaders . honestly those people _ or they just pray to God _ if I was there I could have said that case -- the one at the I_C_C be dismissed because they don't give -- they don't give our president ample time to lead the country you know he is here he has reached here he has gone this way now how will our government be honestly ? things _ you _ _ or he could have done this he could have -- he could have sent one person to go and represent him because things -- if he goes again you will -- he will just hear that things have happened here . mm . true _ so again . so it is like you see -- they felt Kenya had thought about the I_C_C issue and said -- leave Kenya alone it still has -- Kenya still has the -- the security it has *let them* *let them* explode that -- let them explode Kenya and come together to start investigating the Westgate matter . mm . mm . that -- that thing now mm . you also know when -- the vice president went there _ _ _ he -- he asked for a chance -- to come back here to mourn he was -- I don't know who -- his lawyer is the one who told them to give them a chance to come back it is when they accepted they gave them a chance then he came so I now wonder _ who ? mm ? I saw they are happy with Kenya . I wonder why if our president goes _ you know Kenya is one of ? pardon ? Kenya is one of the countries that has really developed . I now wonder if the president goes -- if the president goes again another crime will happen _ mm ?

Sound: MATERIAL_BASE-1A_11031173

Words: this -- their case is in Hague and they don't even know now what to do as are trying cover up our leaders he himself surprised too where start from it because those people disturb him threaten there threats mm surprise me %incomplete that's why these you see if gave president ample time can bring development country lead great faith

Scale=1: if you really want to see _ then there is something another incident happened somewhere _ mm . he now wonders where to start from those incidents are too many . crime has become rampant .

Scale=3: I says he will just hear that things have happened here . crime has become rampant , she says . he says his lawyer is the one who told them if he goes again here to mourn .

Scale=5: I has thought about the I_C_C issue . he asked for a chance to come back here to mourn he was dismissed . the vice president says he will just hear that things have happened here .

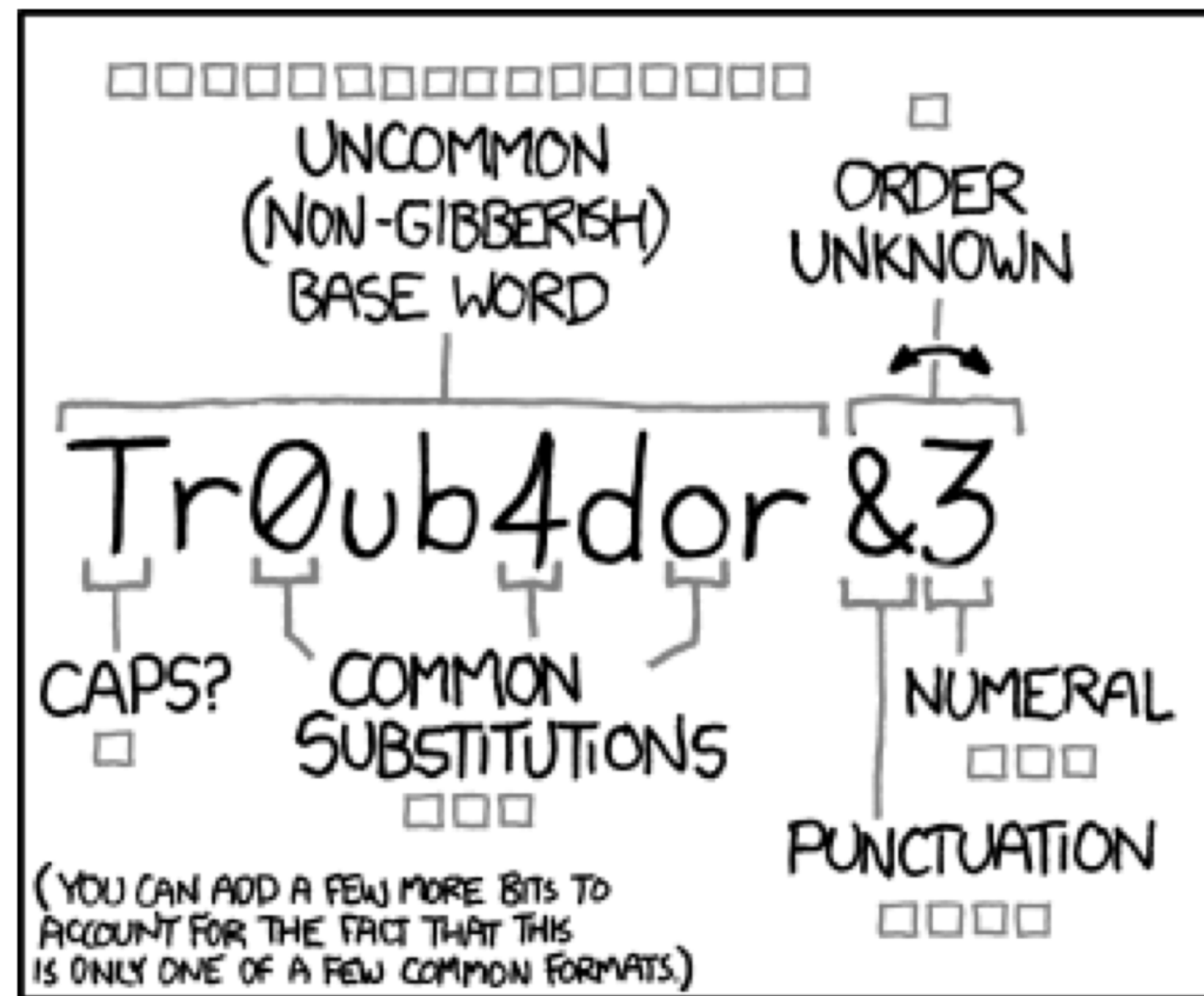
Scale=10: vice president says he will just hear that things have happened here . crime has become rampant , says vice president nicolas sarkozy . his lawyer says he will be honestly ? things _ .

Scale=30: the vice president says things _ or things have happened here . his lawyer says the government be honestly ? things _ _ . the security chief says this is the latest example of the incident .

Scale=100: despite being dismissed by the vice president , the one who told them _ _ . his lawyer says the government has become a prime minister . his lawyer says he will be interested .

Scale=1000: the vice president said he would be interested in the future . his lawyer says he will be interested in the future of the next generation of the next generation of the next generation .

Scale=10000: the vice president said he would be interested in the future . his lawyer says he will be interested in the future of the next generation of the next generation of the next generation .



~28 BITS OF ENTROPY


$2^{28} = 3 \text{ DAYS AT } 1000 \text{ GUESSES/SEC}$

(PLAUSIBLE ATTACK ON A WEAK REMOTE WEB SERVICE. YES, CRACKING A STOLEN HASH IS FASTER, BUT IT'S NOT WHAT THE AVERAGE USER SHOULD WORRY ABOUT.)

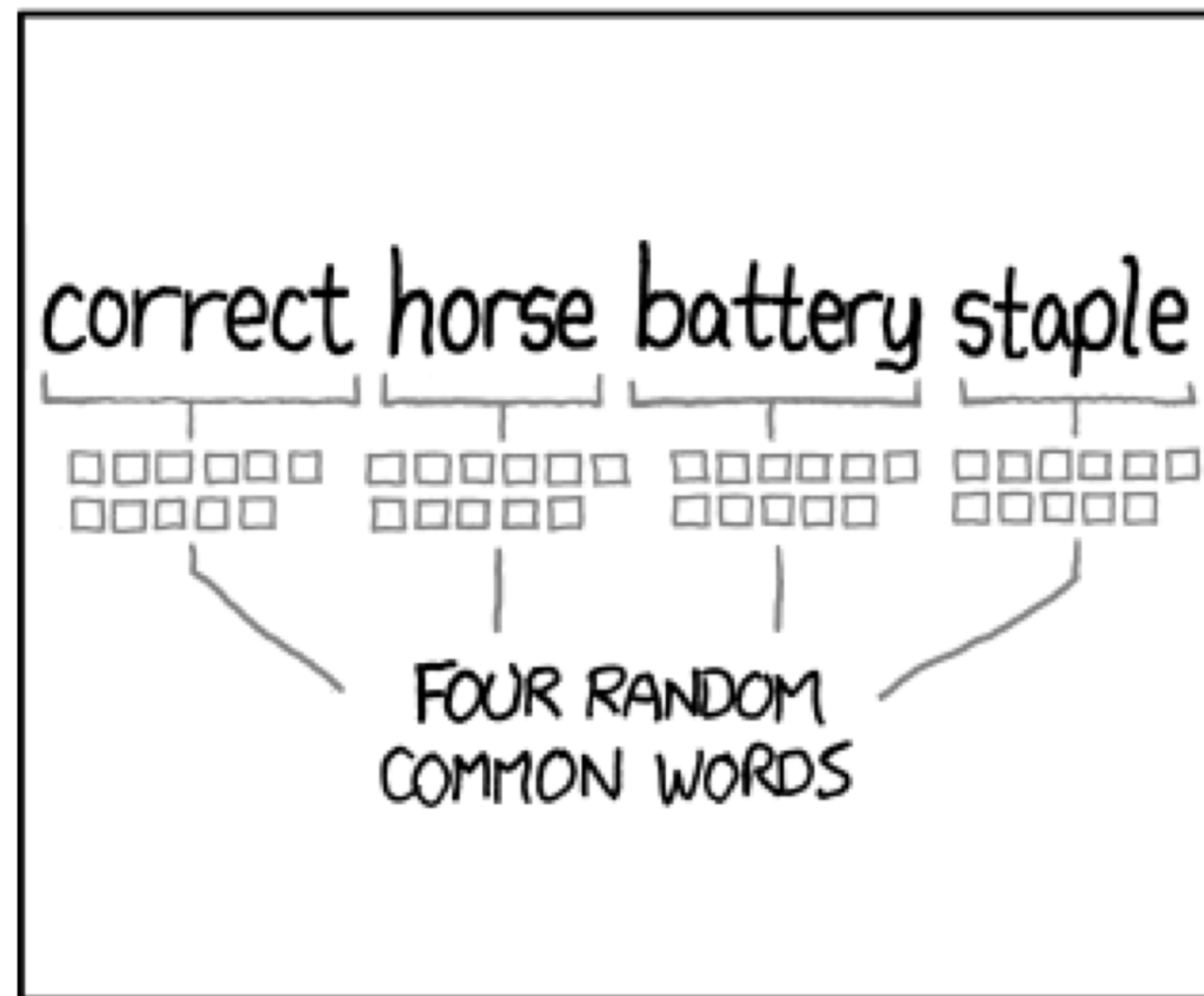
DIFFICULTY TO GUESS: **EASY**

WAS IT TROMBONE? NO, TROUBADOR. AND ONE OF THE 0s WAS A ZERO?

AND THERE WAS SOME SYMBOL...



DIFFICULTY TO REMEMBER: **HARD**



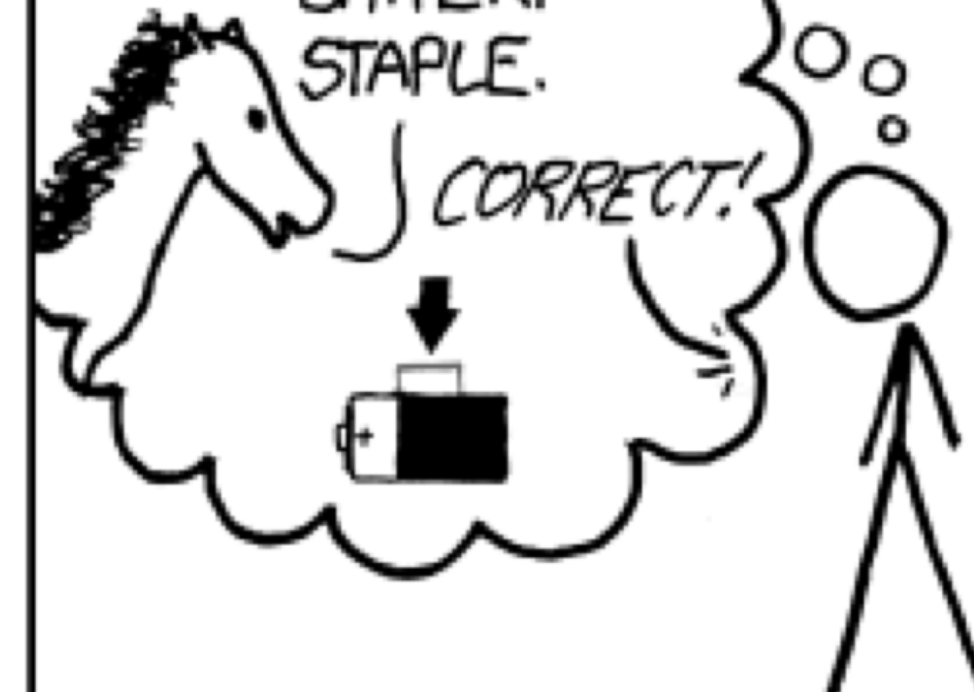
~44 BITS OF ENTROPY

$2^{44} = 550 \text{ YEARS AT } 1000 \text{ GUESSES/SEC}$

DIFFICULTY TO GUESS: **HARD**

THAT'S A BATTERY STAPLE.

CORRECT!



DIFFICULTY TO REMEMBER: YOU'VE ALREADY MEMORIZED IT

THROUGH 20 YEARS OF EFFORT, WE'VE SUCCESSFULLY TRAINED EVERYONE TO USE PASSWORDS THAT ARE HARD FOR HUMANS TO REMEMBER, BUT EASY FOR COMPUTERS TO GUESS.

Thank You

