Accented Speech Recognition with Accent-specific Codebooks

Darshan Prabhu★, Preethi Jyothi★, Sriram Ganapathy✚ and Vinit Unni★
★IIT Bombay and ✚IISc Bangalore

1 What is the central idea?

What?

Handling accents in ASR

Architectural modification to better **handle accent** variation in ASR.

Why?

High variance due to accents

India alone has more than **19,500** languages or dialects.

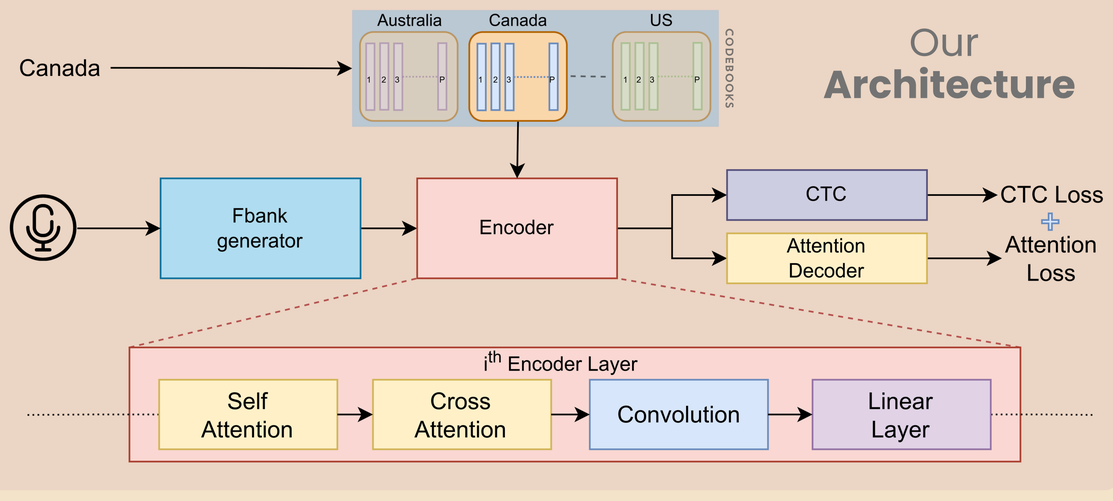
How?

Codebooks and Cross Attention

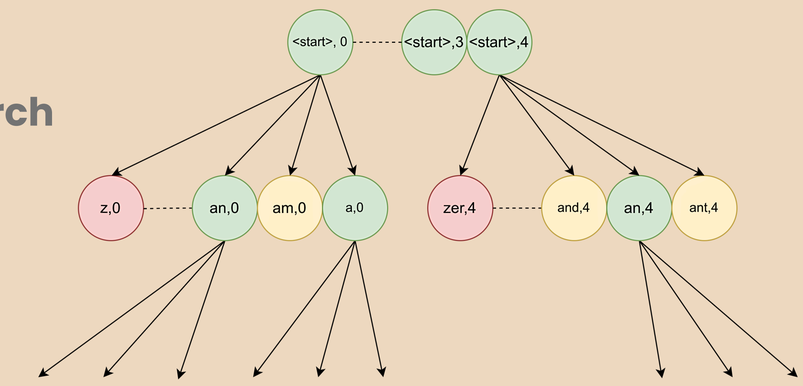
1. Introduce **accent specific information** using codebooks and cross attention.
2. Modify **beam search** to handle accents

2 What are we proposing?

Our Architecture



Joint Beam Search

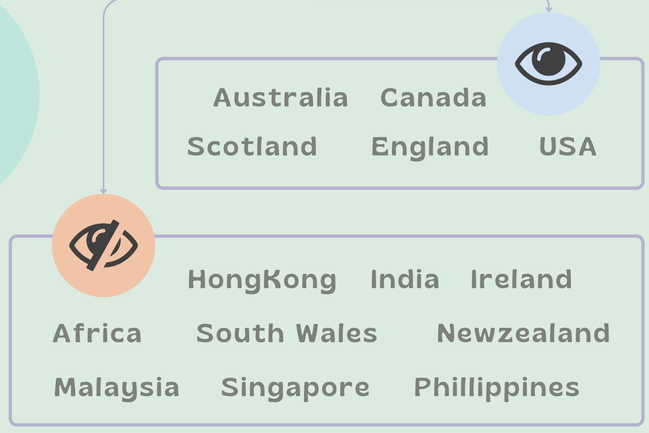


Can ASR be made Accent Aware?

3 About the Dataset

Common Voice

moz://a



5 Future Work

1. **CODEBOOK SIZE**

   Codebook size is a **hyperparameter** that needs to be **finetuned** for each task.

2. **CODEBOOK EXPLOSION**

   Employing accent-specific **codebooks**, one for each accent is **expensive**.

3. **MULTI-ACCENTED SEARCH**

   Our joint beam search allows for each utterance at test time to **commit** to a **single** seen **accent**.

4 Key Results

METHOD	OVERALL	ACCENTS					
		ARABIC	HINDI	KOREAN	MANDARIN	SPANISH	VIATNAMESE
CONFORMER	33.3	30.4	30.4	26.9	37.9	30.3	43.5
I-VECTOR	33.6	31.0	31.2	27.2	38.0	30.4	43.9
MTL	33.4	30.4	30.6	26.9	38.7	30.1	43.7
DAT	33.5	30.7	30.8	26.8	38.3	30.1	43.9
OURS	**32.6**	**29.5**	**30.4**	**26.2**	**37.1**	**29.3**	**42.8**

THE INTERESTING FINDINGS

CODEBOOKS ARE MORE **USEFUL** AT **LOWER** ENCODER LAYERS

HAVING **TOO SMALL** OR **TOO LARGE** NUMBER OF ENTRIES IN CODEBOOK IS **BAD**

FROZEN CODEBOOKS PERFORM **EQUALLY** WELL

PERFORMS **EQUALLY WELL** ON **BALANCED** AND **IMBALANCED** DATASETS

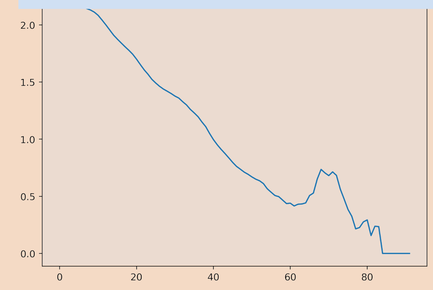
ACCENT USED	OVERALL	SEEN	UNSEEN
TRANSFORMER	22.7	17.3	28.0
CONFORMER	18.9	14.0	23.7
I-VECTOR	18.9	14.1	23.6
MTL	18.9	14.1	23.7
DAT	18.7	14.0	23.4
OURS	**18.2**	**13.6**	**22.9**

Zero Shot Transfer | Information in Codebooks | Ablation Study | Entropy of Beam Search | Overall Results

ACCENT USED	SEEN ACCENTS					UNSEEN ACCENTS								
	AUS	CAN	UK	SCT	US	AFR	HKG	IND	IRL	MAL	NWZ	PHL	SGP	WLS
AUSTRALIA	**11.5**	19.5	17.0	18.1	17.4	22.0	29.8	32.5	24.3	33.7	**18.7**	30.1	37.8	21.1
CANADA	20.5	**14.7**	20.0	15.7	13.5	24.5	27.4	29.6	**21.4**	32.7	25.7	26.6	35.4	21.8
ENGLAND	13.8	17.7	**15.0**	14.4	16.2	**21.5**	27.0	29.9	22.0	**32.3**	21.1	27.1	34.8	**18.0**
SCOTLAND	20.7	17.8	19.1	**10.2**	16.4	24.4	28.2	33.5	22.6	34.4	25.7	29.0	36.6	21.3
US	20.2	**14.7**	19.4	15.5	**13.2**	23.4	**27.0**	**28.1**	21.7	32.4	24.7	**25.8**	**34.3**	22.2

*Numbers reported in this poster are Word Error Rates(WER).

How entropy changes over time?



Poster by: **Darshan Prabhu**

EMNLP 2023

Email: **darshanp@cse.iitb.ac.in**