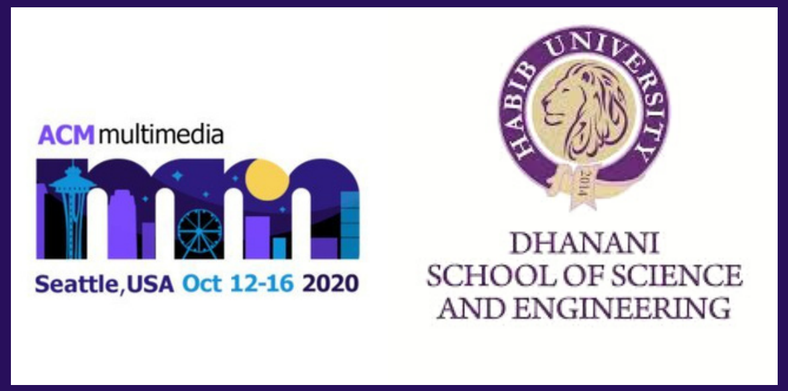
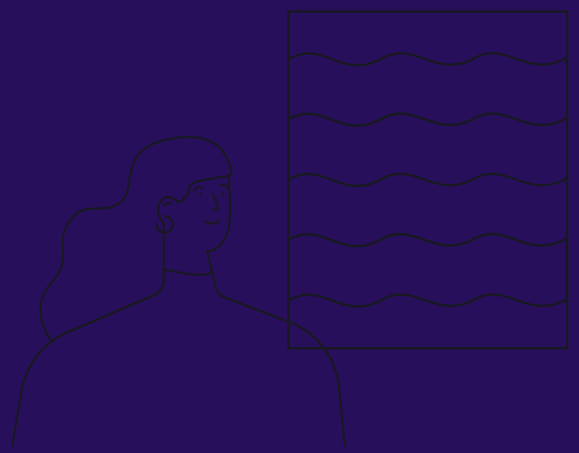


Personalized Audio Stories

Syeda Maryam Fatima, Marina Shehzad, Syed Sami Murtuza, Syeda Saleha Raza

Making Storytelling a Personalized Experience.



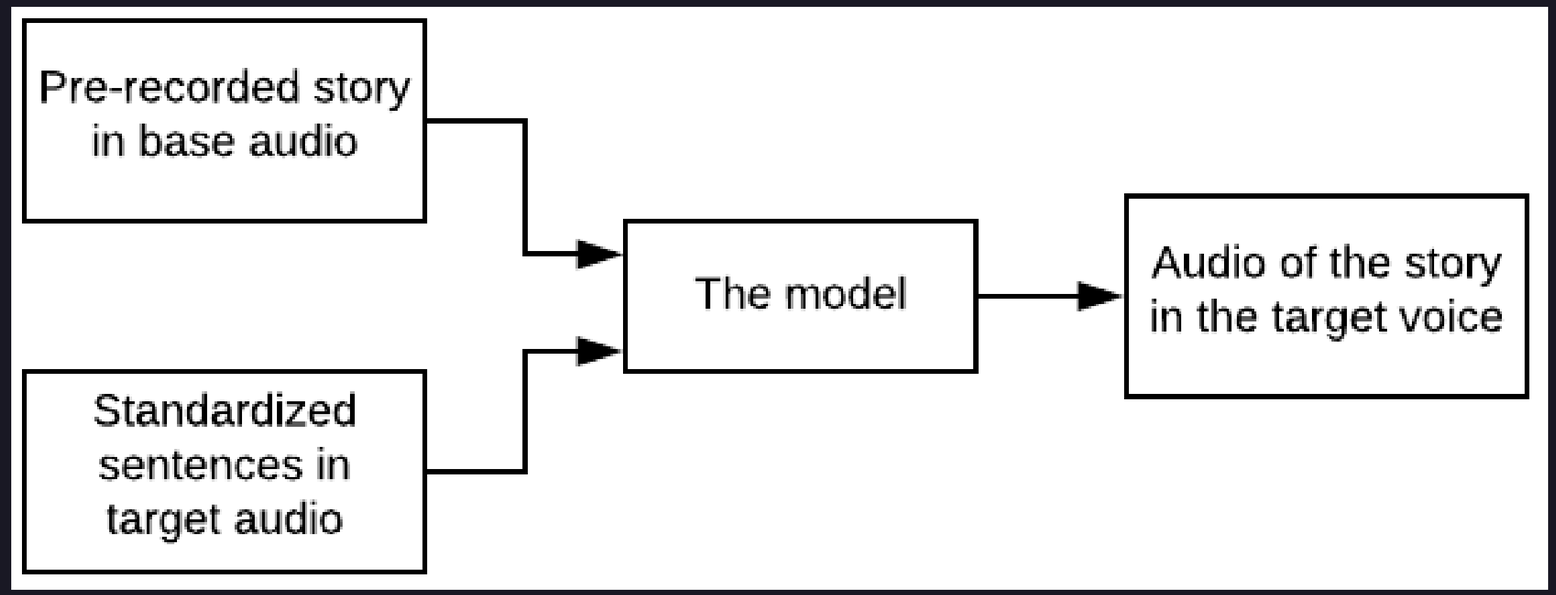
Motivation and Context

Mitigating the disconnect between working parents and their children

Innocuous use of smart devices

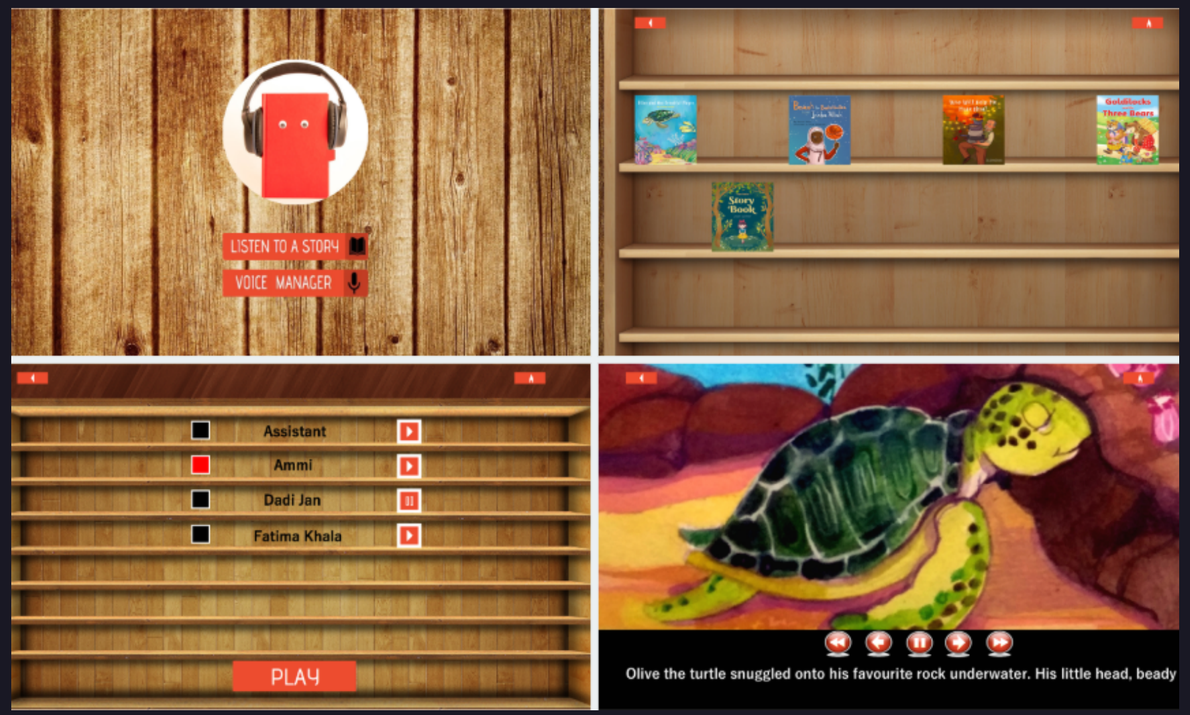
Reviving storytelling culture a

INTRODUCTION



This asks the user to record their voice → Mimics the voice of the parent → This plays the generated audio

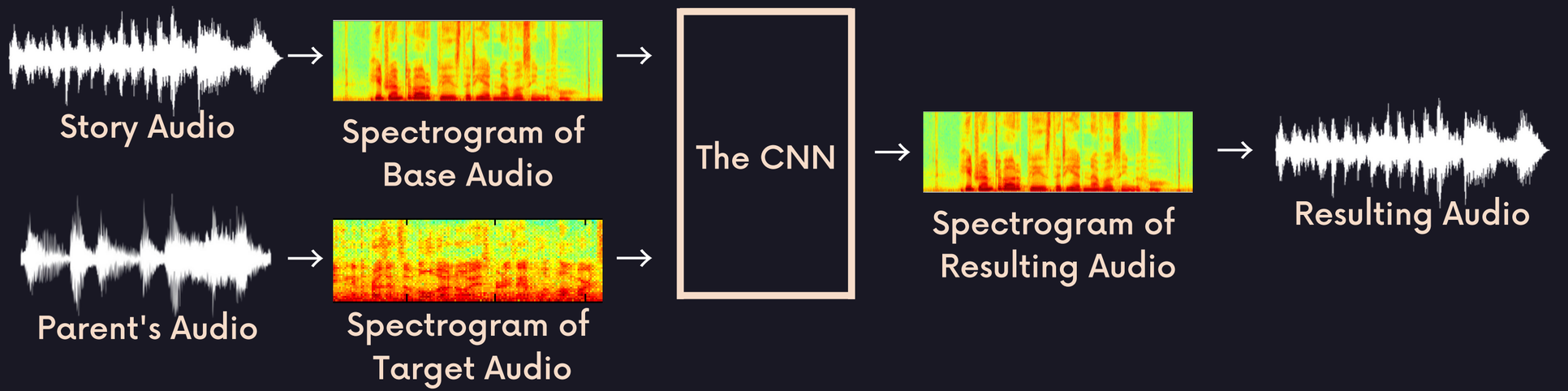
APPLICATION INTERFACE



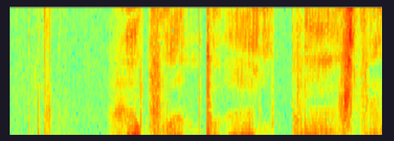
Child's Walkthrough

Parent's Walkthrough

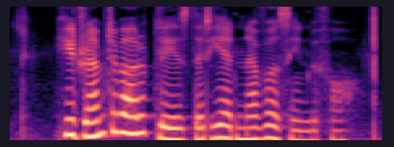
METHODOLGY


Story Audio
Spectrogram of Base Audio
Parent's Audio
Spectrogram of Target Audio
The CNN
Spectrogram of Resulting Audio
Resulting Audio

EXPERIMENTS

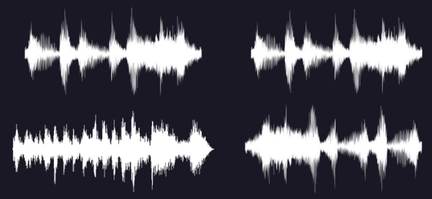


Linear Spectrogram



Mel Spectrogram

Number of Layers



Number of Target Audios

File	Spectrogram	Layer(s)	Base	Target
Lin_1_MS27_SF15	Linear	1	MS-27	SF-15
Lin_2_MS17_SFALL	Linear	2	MS-17	SF-ALL
Lin_2_MS27_SF15	Linear	2	MS-27	SF-15
Mel_1_MS27_SF11	Mel	1	MS-27	SF-11
Mel_2_MS27_FO4	Mel	2	MS-27	F-O-4
Mel_2_MS27_SF15	Mel	2	MS-27	SF-15

DISCUSSION

Gender Difference

Age Difference

Less Time

Small Input Dataset

FUTURE WORK

Languages and accents

Finding Similarity

Applying filters to remove the noise added by CNN

Extending to more domains

Hotline

Video game voice overs

Chatbots

REFERENCES

[1]. A Neural Algorithm of Artistic Style *by* Gatys, Ecker & Bethge
[2]. randomCNN-voice-transfer by mazzzystar (github)
[3]. Speech spectrograms using the fast Fourier transform *by* Oppenheim
[4]Younggun Lee, Taesu Kim, and Soo-Young Lee. 2018. Voice Imitating Text-to-Speech Neural Networks.arXiv preprint arXiv:1806.00927(2018).
[5]Marco Pasini. 2019. Voice Translation and Audio Style Transfer with GANs.Medium(Nov 2019). https://towardsdatascience.com/voice-translation-and-audio-style-transfer-with-gans-b63d58f61854
[6]Hossein Salehghaffari. 2018. Speaker Verification using Convolutional NeuralNetworks.arXiv preprint arXiv:1803.05427(2018).
[7]Hideyuki Tachibana, Katsuya Uenoyama, and Shunsuke Aihara. 2018. Efficientlytrainable text-to-speech system based on deep convolutional networks withguided attention. In2018 IEEE International Conference on Acoustics, Speech andSignal Processing (ICASSP). IEEE, 4784–4788.
[8]Yuxuan Wang, RJ Skerry-Ryan, Daisy Stanton, Yonghui Wu, Ron J Weiss, NavdeepJaitly, Zongheng Yang, Ying Xiao, Zhifeng Chen, Samy Bengio, et al.2017.Tacotron: Towards end-to-end speech synthesis.arXiv preprint arXiv:1703.10135(2017).
[9]Eric Grinstein, Ngoc Duong, Alexey Ozerov, and Patrick Pérez. 2018. Audio styletransfer. https://arxiv.org/abs/1710.11385
[10]WIRED Insider. 2018. How Lyrebird Uses AI to Find Its (Artificial) Voice. https://www.wired.com/brandlab/2018/10/lyrebird-uses-ai-find-artificial-voice/
[11]Ye Jia, Yu Zhang, Ron Weiss, Quan Wang, Jonathan Shen, Fei Ren, Patrick Nguyen,Ruoming Pang, Ignacio Lopez Moreno, Yonghui Wu, et al.2018. Transfer learningfrom speaker verification to multispeaker text-to-speech synthesis. InAdvancesin neural information processing systems. 4480–4490.

THANK YOU