

Can The Automatic Speaker Recognition Detect The Similarity of Siblings Voices?

Amal Al-Zhrani, Mshael Al-Zhrani, Rahaf Al-Thobati, Rawan Al-Zhrani, Wala'a Al-Zhrani.
 Supervisor: Dr. Sarah Al-Shareef
 Computer Sciences Department, Umm Al-Qura University, 2020



Abstract:

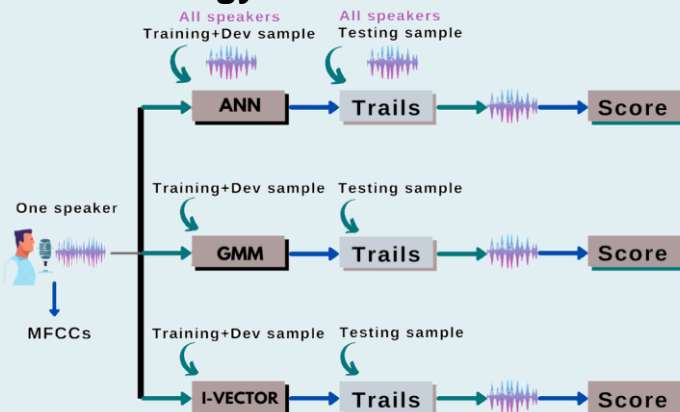
Everyone has a unique voice, because of the different structure of the articulatory organs. Some people have similar voices due to biological similarity or someone is mimicking them. This similarity raises a security concern. Several studies tried to distinguish between similar voices of people using speaker recognition techniques. This study surveys the use of automatic speaker recognition techniques in order to find the similarity between siblings. This includes comparing the performance of several models, such as artificial neural networks (ANNs), Gaussian mixture model (GMM) and i-vector.

Research Objectives:

- Developing siblings dataset (Shiny dataset).
- Compare different models to find the similarity between siblings voices.



Methodology:

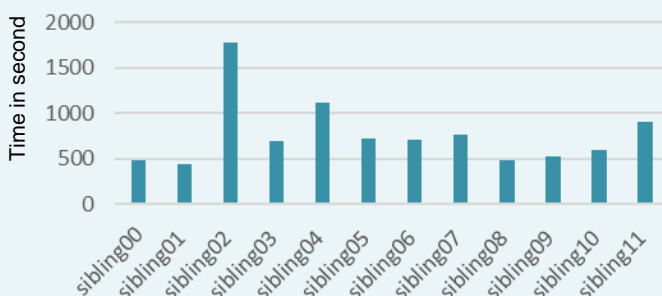


Related Work:

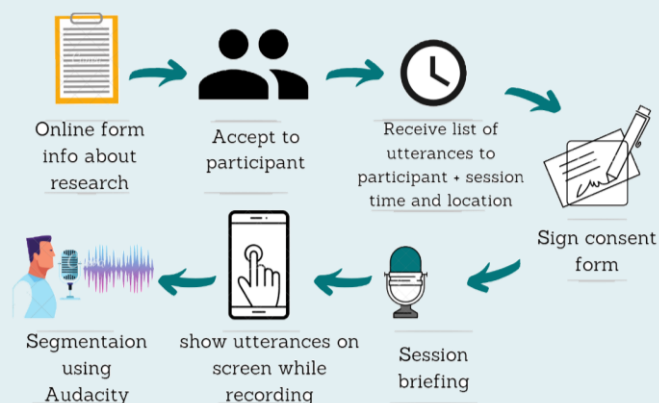
Speaker recognition	The similarity between speakers
Automatic speaker recognition is validating the user's claimed identity using characteristics extracted from their voices.	Two people may sound alike either because of biological similarity or one is mimicking the other
Many research tried to detect the voice similarity between speakers by automatic speaker recognition	
Such as: Kunzel-2010 research, tested if an automatic SPID speaker recognition system can distinguish between identical twins.	

Outcome Dataset:

Total siblings utterances durations



Data Collection:



Development tools:



Results:

Models	EER% :for each experiment		
	One hidden layer	Two hidden layer	Three hidden layer
ANN	24.10%	25.12%	29.11%
GMM	31.25%	21.88%	34.37%
i-vector	12.5%	9.37%	65.62%

Conclusion:

In this study, we explored the similarity between siblings through speaker recognition models. First, we collected, cleaned and processed shiny dataset of 28 speakers with 9229s duration. This dataset was used to build three models: ANN, GMM, and i-vector. Then the models' performance was compared. The i-vector has shown the best performance with EER 9.375%.

Contact:

Group email: cs04.2020@gmail.com
 Supervisor email: saashareef@uqu.edu.sa

E - copy:



References:

