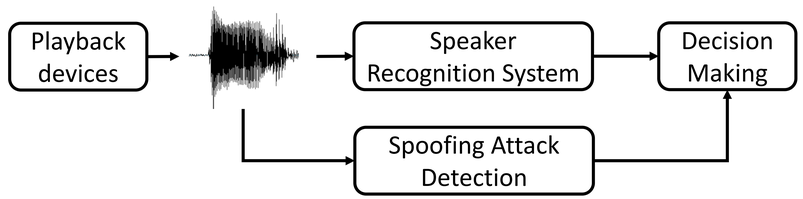
RECURRENT NEURAL NETWORKS FOR AUTOMATIC REPLAY SPOOFING ATTACK DETECTION

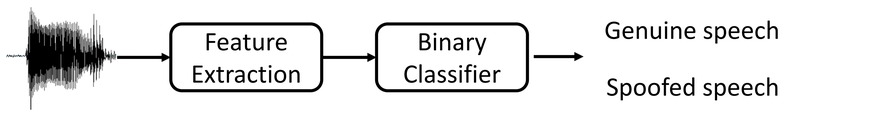
Zhuxin Chen[1], *Weibin Zhang*[1], *Zhifeng Xie*[1], *Xiangmin Xu*[1], *Dongpeng Chen*[2]
[1]South China University of Technology, [2]Shengyang Technology Company

Introduce

➢ Replay Spoofing Attack to Speaker Recognition System

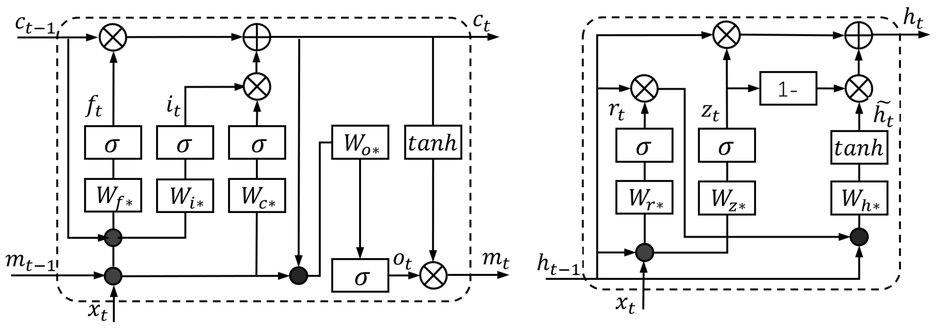


➢ Automatic Replay Spoofing Attack Detection

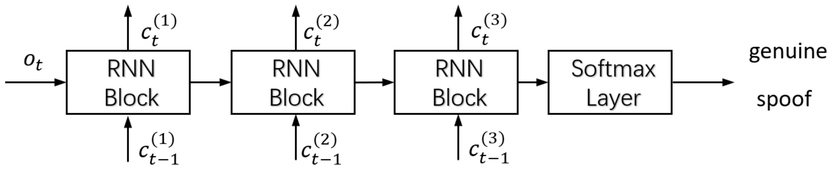


Model

➢ Single Unit of LSTM and GRU



➢ RNN for spoof detection



$$score = \frac{1}{T}\left(\sum_{t=1}^{T} logP(\text{genuine}|o_t) - logP(spoof|o_t)\right)$$

Experiment Setup

➢ Database
ASVspoof2017 and BTAS2016 were used.
Both the evaluation sets contain recordings with unknown replay conditions.

➢ Feature
CQCC: 30 coefficients (achieved the best on ASVspoof2015)
MFCC: 120 filters*, 30 cepstral coefficients
Fbank: 120 filters*
*Increasing to 120 filters significantly improve the recognition accuracy.

➢ Modelling
DNN: 11-frame context window, three hidden layers with 512 units and a softmax layer
RNN: three recurrent layers with 256 cells and a softmax layer, sequence categorical cross-entropy loss function

Experiment Result

Model	DNN						LSTM				GRU			
Feature	CQCC		MFCC		Fbank		MFCC		Fbank		MFCC		Fbank	
Dataset	DEV	EVAL	DEV	EVAL	DEV	EVAL	DEV	EVAL	DEV	EVAL	DEV	EVAL	DEV	EVAL
EER	5.44	20.36	7.59	12.87	8.09	12.13	10.06	14.42	6.88	10.98	10.39	14.18	6.32	**9.81**

Table 1 shows the results on ASVspoof 2017. The MFCC and Fbank features significantly outperform the CQCC feature. the GRU model with Fbank feature achieves the best EER of 9.81%, which outperforms the best feed-forward neural network by 19% relatively.

Model	DNN				LSTM				GRU			
Feature	MFCC		Fbank		MFCC		Fbank		MFCC		Fbank	
Dataset	DEV	EVAL	DEV	EVAL	DEV	EVAL	DEV	EVAL	DEV	EVAL	DEV	EVAL
ALL	1.153	2.058	0.779	2.007	0.273	2.149	0.052	1.107	0.241	1.912	0.039	**1.077**
RE-LP-LP	0.378	0.773	0.234	0.783	0.215	1.102	0.019	0.528	0.192	2.197	0.019	0.443
RE-LP-HQ-LP	2.905	2.534	2.213	2.308	0.425	1.893	0.122	1.182	0.39	2.038	0.114	0.752
RE-PH1-LP	0.266	1.312	0.254	0.662	0.168	0.698	0.056	0.141	0.235	0.897	0.031	0.191
RE-PH2-LP	0.128	1.002	0.065	0.908	0.063	0.568	0.018	0.209	0.055	1.734	0.019	0.267
RE-PH2-PH3	-	2.521	-	2.517	-	2.461	-	0.495	-	2.364	-	0.53
RE-LPPH2-PH3	-	2.622	-	2.994	-	3.717	-	2.32	-	3.184	-	2.592

Table 2 shows the results on BTAS 2016. The best GRU model outperforms the best DNN model by 46% relatively.